

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2005, 11:07:54 ; Search time 15263 Seconds
(without alignments)
11752.684 Million cell updates/sec

Title: US-10-600-862A-6
Perfect score: 3702
Sequence: 1 cttaccgcatgagggtcc.....tcagctggtgactcgag 3702

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues 9416466
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sta.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2647.8	71.5	3624	6	AX840238 Sequence
2	2647.8	71.5	3624	6	Y00839 H. sapiens G
3	2647.8	71.5	3846	6	CQ882052 Sequence
4	2646.2	71.5	3625	6	CQ723359 Sequence
5	2639.8	71.3	3441	9	M34424 Human acid
6	2635	71.2	3687	9	BC040431 Homo sapi
7	2585.8	69.8	3626	11	BV177050 sqnm92407
8	2033.2	54.9	2814	4	AF171665 Bos tauru
9	1903	51.4	3408	10	BC061753 Rattus no
10	1899.8	51.3	3504	10	BC010210 Mus muscu
11	1881.4	50.8	3364	6	AX305670 Sequence
12	1881.4	50.8	3364	10	MMU49351 Mus muscu
13	1332	36.0	3112	5	AB006754 Coturnix
14	994.4	26.9	3548	5	AB000967 Coturnix
15	957	25.9	1493	6	CQ776406 Sequence
16	957	25.9	1493	6	AX052732 Sequence
17	957	25.9	1493	6	AX557300 Sequence
18	957	25.9	1493	9	M63959 Human alpha
19	955.4	25.8	1519	9	BS647984 Homo sapi

20	953.8	25.8	1499	6	BD134425	Human nuc
21	953.8	25.8	1499	6	AX017254	Sequence
22	953.8	25.8	1499	6	AX524958	Sequence
23	895.4	24.2	1485	6	CQ718547	Sequence
24	665.6	18.0	1419	10	BC057979	Mus muscu
25	665.6	18.0	1451	10	BC032170	Mus muscu
26	665.6	18.0	1478	10	S67967	heparin bin
27	665.6	18.0	2902	6	CQ777321	Sequence
28	665.6	18.0	2902	6	AX305444	Sequence
29	665.6	18.0	2902	6	MUSHBP44	Mus musculu
30	665.6	18.0	2913	10	BC046641	Mus muscu
31	665.6	18.0	1373	10	BC082020	Rattus no
32	664	17.9	1452	10	AJ639614	Mus muscu
33	663.8	17.9	2768	10	RN45KDA	Zi1994 R.norvegicu
34	616.6	16.7	2753	10	BC059887	Mus muscu
35	581.6	15.7	1260	10	RATGP330	M31051 Rat Heymann
36	509	13.7	1624	6	BD134483	Human nuc
37	509	13.7	1624	6	AX017383	Sequence
38	509	13.7	1624	6	AX525087	Sequence
39	465.8	12.6	1497	5	GDAJ1912	AJ001912 Gallus do
40	424.4	11.5	969	11	BV179107	BV179107 sqnm10389
41	415.4	11.2	522	11	BV168003	sqnm7017
42	388.6	10.5	2873	8	AK105449	Oryza sat
43	387.2	10.5	1853	5	BC054293	Xenopus 1
44	382.8	10.3	2804	8	AK121588	Oryza sat
45	371.2	10.0	6513	9	AF016833	Homo sapi

ALIGNMENTS

RESULT 1
AX840238
LOCUS AX840238 3624 bp DNA linear PAT 16-DEC-2003
DEFINITION Sequence 12 from Patent WO03073839.
ACCESSION AX840238
VERSION AX840238.1 GI:39978638
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Fogher, C. and Reggi, S.
TITLE Expression of lysosomal enzymes in plant seeds
JOURNAL Patent: WO 03073839-A 12 12-SEP-2003;
FEATURES
Location/Qualifiers
1..3624
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
sig_peptide 220..426
CDS 427..3078
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE85448.1"
/db_xref="GI:39978639"
/translation="AHPGRPRAPVPTQCDVPPNSRFDCAPKAITQBCARGCCYTPA
KQGLQAGQWQWCFPPSPYSYKLENLSSEMGYTATLTRTPPKDILTLRLDV
MMETENRLHFTIKDPANRYEVEPLETPRVHSRAPSPLYSVFSEEFVIVHRLQDGR
VLLNTTAPLFFADQLQSLSPQYITGLAEHLSPLMLSTWTRITLWNRDLATPP
GANLYGSHPPFYLALEDGSAHGVLINNSAMDVLPQSPALSWRSTGGILDVFLQGP
EPKSVVQYLDVVGVPFMPYMGVGLHLCRMGYSSTAITRQVVNTRAHFPLDVQWN
DLDDWDSRDTFNKGDPRDPAMQELHOGGRRYMMI VDPATSSSGPAGSVRYPDEG
LRRGFINTETQPLIGKWPSTAPDFTNPTALAWMEDMAEFHQVDFDGMWIDM
NPSFPIRGSGCCNNELNPPYFVGVGGILQATTCASHQFSLTHNLNLVGL
TEAISHALVARGTRPFVSRSTFAGRYAGHWGVDVWSWGLQASVSPILQFN
LIGVLVGVADVCQFLGNTSEELVMTQLGAFYPPMRNNSLSLSPQYSEPAQO
AMRKALTRYALLPHLYTLFHOAHVAGETVARPLFPPKSDSTTQVDFQLLMGEALL
ITPVLQAGKAEVTGYPLGTVDLOTVPTEALGSLPPPPAPAPREPAIHSEGWVTLPA
PLDTINHLRAGYIIPLOGFLTTTESRQOPPALAVLTGGEARGELFDDGESLEV
LERGATQVIFLARNTIINVELRVTSSEGAGLQQLQKVTVLGVATAPQQVLSNVPVSN
FTYSPDTRKVLIDICVSLLMGEQLVSNWC"

ORIGIN									
Query Match 71.5%; Score 2647.8; DB 6; Length 3624;									
Best Local Similarity 99.7%; Pred. No. 0;									
Matches 2652; Conservative 0; Mismatches 7; Indels 0; Gaps 0;									
QY	1036	GAACCGGTGACACACCCCGCGCTCCAGAGAGTCCACACAGTGCAGACGTCCCGCCC	1095						
DB	418	GATGCCAGGCACACCCCGCGCTCCAGAGAGTGCACACAGTGCAGACGTCCCGCCC	477						
QY	1096	AACAGCCGTTGATGCGCCCTGACAAAGGCCATCACCCAGAAACAGTGCAGAGCCCGC	1155						
DB	478	AACAGCCGTTGATGCGCCCTGACAAAGGCCATCACCCAGAAACAGTGCAGAGCCCGC	537						
QY	1156	GGCTGCTGTACATCCTCGCAAGCAGGGCTGACAGGAGCCAGATGGGAGCCCTGG	1215						
DB	538	GGCTGCTGTACATCCTCGCAAGCAGGGCTGACAGGAGCCAGATGGGAGCCCTGG	597						
QY	1216	TGCTTCTTCCACCCAGCTACCCAGCTACAAAGCTGGAGAACCTGAGCTCCTCTGAATG	1275						
DB	598	TGCTTCTTCCACCCAGCTACCCAGCTACAAAGCTGGAGAACCTGAGCTCCTCTGAATG	657						
QY	1276	GGCTACACGGCCACCTGACCCCGTACACCCCGACCTTCTTCCCAAGACATFCTGACC	1335						
DB	658	GGCTACACGGCCACCTGACCCCGTACACCCCGACCTTCTTCCCAAGACATFCTGACC	717						
QY	1336	CTGCGCTGAGCGTGTGATGGAGCTGAGACCGGCTCCACTTACAGATCAAGATCCA	1395						
DB	718	CTGCGCTGAGCGTGTGATGGAGCTGAGACCGGCTCCACTTACAGATCAAGATCCA	777						
QY	1396	GCTAACAGGCGCTACAGAGTGCCCTTGGAGACCCCGCTGTCCACAGCCGGGACCGTCC	1455						
DB	778	GCTAACAGGCGCTACAGAGTGCCCTTGGAGACCCCGCTGTCCACAGCCGGGACCGTCC	837						
QY	1456	CCACTCTACAGCGTGAAGTCTCCAGAGAGCCCTTCGGGGTATGTGACCCGAGCTG	1515						
DB	838	CCACTCTACAGCGTGAAGTCTCCAGAGAGCCCTTCGGGGTATGTGACCCGAGCTG	897						
QY	1516	GACGGCCGGTGTGTGACACGAGCGTGGGCGCCCTGTTCTTGGGACCAAGTTCCTT	1575						
DB	898	GACGGCCGGTGTGTGACACGAGCGTGGGCGCCCTGTTCTTGGGACCAAGTTCCTT	957						
QY	1576	CAGCTGTCAACCTTCGCTCGCAGTATATACAGAGCTCGCCGAGCACCTCAGTCCC	1635						
DB	958	CAGCTGTCAACCTTCGCTCGCAGTATATACAGAGCTCGCCGAGCACCTCAGTCCC	1017						
QY	1636	CTGATGCTCAGACACAGCTGAGACAGGATCACCTGTGAAACCGGAGCCTTGGCCCAAG	1695						
DB	1018	CTGATGCTCAGACACAGCTGAGACAGGATCACCTGTGAAACCGGAGCCTTGGCCCAAG	1077						
QY	1696	CCCGGTGCGAACCTCTACGGGTCTACCTTTCTACCTGCGCTGAGGAGCGCGGTGCG	1755						
DB	1078	CCCGGTGCGAACCTCTACGGGTCTACCTTTCTACCTGCGCTGAGGAGCGCGGTGCG	1137						
QY	1756	GCACACGGGGTGTCTCTGCTAAACAGCAATGCCATGATGTGTCTGAGCGCGAGCCCT	1815						
DB	1138	GCACACGGGGTGTCTCTGCTAAACAGCAATGCCATGATGTGTCTGAGCGAGCCCT	1197						
QY	1816	GGCTTATGCTGAGGTGACAGGTGGATCTCTGGATGTCTACATCTTCTGGGCCAGAG	1875						
DB	1198	GGCTTATGCTGAGGTGACAGGTGGATCTCTGGATGTCTACATCTTCTGGGCCAGAG	1257						
QY	1876	CCCAAGAGCGTGTGACAGTACCTGAGCGTGTGGATACCCGTTTCATGCGGCATAC	1935						
DB	1258	CCCAAGAGCGTGTGACAGTACCTGAGCGTGTGGATACCCGTTTCATGCGGCATAC	1317						
QY	1936	TGGGGCTTGGGCTTCCACCTGTGCGCTGGGGTACTCTCTCACCCGCTATCACCCGCCAG	1995						
DB	1318	TGGGGCTTGGGCTTCCACCTGTGCGCTGGGGTACTCTCTCACCCGCTATCACCCGCCAG	1377						
QY	1996	GTGGTGGAGAACATGACAGAGGCCCACTTCCCGCTGGACGTGCAATGGAAACGACTGGAC	2055						
DB	1378	GTGGTGGAGAACATGACAGAGGCCCACTTCCCGCTGGACGTGCAATGGAAACGACTGGAC	1437						
QY	2056	TACATGAGACTCCCGAGGAGACTTACGTTCAACAGGATGGCTTCGGGACTTCCGGCC	2115						
DB	1438	TACATGAGACTCCCGAGGAGACTTACGTTCAACAGGATGGCTTCGGGACTTCCGGCC	1497						
QY	2116	ATGGTGCAGGAGCTGCACACAGGCGCCGGCGCTACATGATGATCGTGGATCTTGCATC	2175						
DB	1498	ATGGTGCAGGAGCTGCACACAGGCGCCGGCGCTACATGATGATCGTGGATCTTGCATC	1557						
QY	2176	AGCAGCTCGGGCCCTTCGGGAGCTACAGGCCCTACGAGGAGGTTCGGGAGGGGGTT	2235						
DB	1558	AGCAGCTCGGGCCCTTCGGGAGCTACAGGCCCTACGAGGAGGTTCGGGAGGGGGTT	1617						
QY	2236	TTTCATCACCAACGAGACCGGCAGCGCTGATTTGGGAGGTATGGCCCGGGTCCAATGCC	2295						
DB	1618	TTTCATCACCAACGAGACCGGCAGCGCTGATTTGGGAGGTATGGCCCGGGTCCAATGCC	1677						
QY	2296	TTTCCCGACTTTCACAAACCCCAACAGCCCTGGCTCGTGGAGGACATGATGCTGATTC	2355						
DB	1678	TTTCCCGACTTTCACAAACCCCAACAGCCCTGGCTCGTGGAGGACATGATGCTGATTC	1737						
QY	2356	CATGACCAAGTGCCCTTCGACGGCTTGTGGATTGACATGAAAGAGCCTTCCAATTCATC	2415						
DB	1738	CATGACCAAGTGCCCTTCGACGGCATGTGGATTGACATGAAAGAGCCTTCCAATTCATC	1797						
QY	2416	AGAGGCTCTGAGGACGGCTGCCCAACATGAGCTGGAGAACCCACCTTACGCTGGG	2475						
DB	1798	AGAGGCTCTGAGGACGGCTGCCCAACATGAGCTGGAGAACCCACCTTACGCTGGG	1857						
QY	2476	GTGGTTGGGGGACCTCCAGCGGCCACCAATCTGTGCTCCAGCCACACAGTTTCTCTCC	2535						
DB	1858	GTGGTTGGGGGACCTCCAGCGGCCACCAATCTGTGCTCCAGCCACACAGTTTCTCTCC	1917						
QY	2536	ACACACTACACCTTCGACCAACCTCTACGGCTTGACCGAAGCCATCGCTCCACAGGGG	2595						
DB	1918	ACACACTACACCTTCGACCAACCTCTACGGCTTGACCGAAGCCATCGCTCCACAGGGG	1977						
QY	2596	CTGGTGAAGGCTCGGGGACACCGCCATTGTTGATCTCCCGCTCGACCTTTCTCTGGCCAC	2655						
DB	1978	CTGGTGAAGGCTCGGGGACACCGCCATTGTTGATCTCCCGCTCGACCTTTCTCTGGCCAC	2037						
QY	2656	GGCCGATACGCGGCCACTGGAAGGGGACGTTGTGAGCTCTCTGGAGCAGCTCGCTCC	2715						
DB	2038	GGCCGATACGCGGCCACTGGAAGGGGACGTTGTGAGCTCTCTGGAGCAGCTCGCTCC	2097						
QY	2716	TCCGTGCCAGAACTCTCGAGTTTAACTGTCTGGGGGTGCTCTGTTGGGGCCGACGTC	2775						
DB	2098	TCCGTGCCAGAACTCTCGAGTTTAACTGTCTGGGGGTGCTCTGTTGGGGCCGACGTC	2157						
QY	2776	TGCGGCTTCTGGGCAACACCTCAGAGGAGCTGTGTGCTGCTGGACCCAGCTGGGGGCC	2835						
DB	2158	TGCGGCTTCTGGGCAACACCTCAGAGGAGCTGTGTGCTGCTGGACCCAGCTGGGGGCC	2217						
QY	2836	TTCTACCCCTTCATCGGAAACCAACAGCTGTCTAGTCTGCCCCAGAGCCGTACAGC	2895						
DB	2218	TTCTACCCCTTCATCGGAAACCAACAGCTGTCTAGTCTGCCCCAGAGCCGTACAGC	2277						
QY	2896	TTACAGGAGCCGCCACGACGACATGAGGAGCCCTCACCTCGGCTTACGACCTCTC	2955						
DB	2278	TTACAGGAGCCGCCACGACGACATGAGGAGCCCTCACCTCGGCTTACGACCTCTC	2337						
QY	2956	CCCCACCTTACACACTGTTCACAGGCCACGCTCGGGGGGAGACCGTGGCCCGGGCC	3015						
DB	2338	CCCCACCTTACACACTGTTCACAGGCCACGCTCGGGGGGAGACCGTGGCCCGGGCC	2397						
QY	3016	CTCTTCTGGAGTTCCTCAAGGACTCTAGCACTGTGGAACCAACAGCTCTGTGG	3075						
DB	2398	CTCTTCTGGAGTTCCTCAAGGACTCTAGCACTGTGGAACCAACAGCTCTGTGG	2457						
QY	3076	GGGGAGGCCCTGCTCATACCCAGTGTCTCAGGGCCGGGAGCCGAGTGTACTGGCTAC	3135						
DB	2458	GGGGAGGCCCTGCTCATACCCAGTGTCTCAGGGCCGGGAGCCGAGTGTACTGGCTAC	2517						

[illegible]

Db 1360 GCACACGGGGTGTCTCTGCTAAACAGCAATGCCATGTGATGTGGTCTCTGAGCGCCAGCCCT 1419
QY 1816 GCCCTTAGCTGGAGTTCGACAGGTGGGATCCCTGGATGTCTACATCTTCTCTGGGCCAGAG 1875
Db 1420 GCCCTTAGCTGGAGTTCGACAGGTGGGATCCCTGGATGTCTACATCTTCTCTGGGCCAGAG 1479
QY 1876 CCCAAGAGCGTGGTGCAGCAGTACCTTGGACGTTGTGGATACCCGTTTCATGCGGCATAC 1935
Db 1480 CCCAAGAGCGTGGTGCAGCAGTACCTTGGACGTTGTGGATACCCGTTTCATGCGGCATAC 1539
QY 1936 TGGGGCTTGGCTTCCACCTGTGCGCTGGGGTACTCTCCACCGCTATACCGCGCCAG 1995
Db 1540 TGGGGCTTGGCTTCCACCTGTGCGCTGGGGTACTCTCCACCGCTATACCGCGCCAG 1599
QY 1996 GTGGTGGAGAACATGACACAGGGCCCACTTCCCGCTGGAGCTCAATGGAAACGACCTGGAC 2055
Db 1600 GTGGTGGAGAACATGACACAGGGCCCACTTCCCGCTGGAGCTCAATGGAAACGACCTGGAC 1659
QY 2056 TACATGGACTCCCGAGGAGCTTTCAGGTTCAACAGGATGGCTTCCGGAGCTTCCCGGCC 2115
Db 1660 TACATGGACTCCCGAGGAGCTTTCAGGTTCAACAGGATGGCTTCCGGAGCTTCCCGGCC 1719
QY 2116 ATGGTGCAGAGCTGCACACAGGGCGCGCGCTACATGATGATCGTGGATCCTGCCATC 2175
Db 1720 ATGGTGCAGAGCTGCACACAGGGCGCGCGCTACATGATGATCGTGGATCCTGCCATC 1779
QY 2176 AGCAGCTCGGGCCCTGCGGGAGCTACAGGCCCTACGACAGGGTCTGCGAGGGGGTT 2235
Db 1780 AGCAGCTCGGGCCCTGCGGGAGCTACAGGCCCTACGACAGGGTCTGCGAGGGGGTT 1839
QY 2236 TTTCATCAACAGAGACCGGCCAGCGCTTGAATGGGAAGGTATGGCCCGGGTCCATGCCC 2295
Db 1840 TTTCATCAACAGAGACCGGCCAGCGCTTGAATGGGAAGGTATGGCCCGGGTCCATGCCC 1899
QY 2296 TTCCCCGACTTACCAACCCACAGCCCTGGCTGTGGAGGACATGGTGTGCTGATTC 2355
Db 1900 TTCCCCGACTTACCAACCCACAGCCCTGGCTGTGGAGGACATGGTGTGCTGATTC 1959
QY 2356 CATGACAGGTGCTTTCGACGGCTGTGGATTTGACATGAACAGAGCTTCCAACTTCATC 2415
Db 1960 CATGACAGGTGCTTTCGACGGCTGTGGATTTGACATGAACAGAGCTTCCAACTTCATC 2019
QY 2416 AGAGGCTCTGAGACGGCTGCCCAACATAGCTGGAGAACCCACCTTACGTGCTGGG 2475
Db 2020 AGAGGCTCTGAGACGGCTGCCCAACATAGCTGGAGAACCCACCTTACGTGCTGGG 2079
QY 2476 GTGGTTGGGGGACCTCCAGCGGCCACCATCTGTGCTCCAGCCACAGTTTCTCTCC 2535
Db 2080 GTGGTTGGGGGACCTCCAGCGGCCACCATCTGTGCTCCAGCCACAGTTTCTCTCC 2139
QY 2536 ACACACTCAACCTGCACAACTCTACCGCCCTGACCGAAGCCATCGCCTCCACAGGGCG 2595
Db 2140 ACACACTCAACCTGCACAACTCTACCGCCCTGACCGAAGCCATCGCCTCCACAGGGCG 2199
QY 2596 CTGGTGAAGGCTCGGGGACAGCCCATTTGTGATCTCCCGCTGACCTTTGCTGGCCAC 2655
Db 2200 CTGGTGAAGGCTCGGGGACAGCCCATTTGTGATCTCCCGCTGACCTTTGCTGGCCAC 2259
QY 2656 GCGCGATACGCGGCCACTGAGCGGGGACGTGTGGAGTCTCTGGAGCAGCTCGCCTCC 2715
Db 2260 GCGCGATACGCGGCCACTGAGCGGGGACGTGTGGAGTCTCTGGAGCAGCTCGCCTCC 2319
QY 2716 TCCGTGCCAGAACTCTGCAAGTTTAACTCTGCTGGGGTCCCTCTGCTGGGGCCCGACGTC 2775
Db 2320 TCCGTGCCAGAACTCTGCAAGTTTAACTCTGCTGGGGTCCCTCTGCTGGGGCCCGACGTC 2379
QY 2776 TCGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGGCTGAGCCACGAGCTGGGGCC 2835
Db 2380 TCGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGGCTGAGCCACGAGCTGGGGCC 2439
QY 2836 TTCTACCCCTTCATGCGGAAACCAACAGCCTGCTCAGTCTGCTCCCGAGGCGGTACAGC 2895
Db 2440 TTCTACCCCTTCATGCGGAAACCAACAGCCTGCTCAGTCTGCTCCCGAGGCGGTACAGC 2499

QY 2896 TTTCAGCAGCGGGCCAGCAGGCCATGAGGAAGGCCCTCACCTCGCTACGCACTCTCTC 2955
Db 2500 TTTCAGCAGCGGGCCAGCAGGCCATGAGGAAGGCCCTCACCTCGCTACGCACTCTCTC 2559
QY 2956 CCCCACTCTACACACTGTTTCCACAGGCCACGCTGCGGGGGAGACCGTGGCCGGGCC 3015
Db 2560 CCCCACTCTACACACTGTTTCCACAGGCCACGCTGCGGGGGAGACCGTGGCCGGGCC 2619
QY 3016 CTCTTCTTGGAGTTCCTCCCAAGGACTCTAGCACTCTGAGACTGTGGACACACAGCTCTGTGG 3075
Db 2620 CTCTTCTTGGAGTTCCTCCCAAGGACTCTAGCACTCTGAGACTGTGGACACACAGCTCTGTGG 2679
QY 3076 GGGGAGGCGCTGCTCATCACCCAGTGTCCAGGCCGGGAAGGCCGAAGTGAATGCTGCTAC 3135
Db 2680 GGGGAGGCGCTGCTCATCACCCAGTGTCCAGGCCGGGAAGGCCGAAGTGAATGCTGCTAC 2739
QY 3136 TTCCCCCTTGGGACATGTAAGACCTGACAGCGGTGCCAATAGAGGCCCTTGGGAGCCTC 3195
Db 2740 TTCCCCCTTGGGACATGTAAGACCTGACAGCGGTGCCAATAGAGGCCCTTGGGAGCCTC 2799
QY 3196 CCACCCCCACCTGACGCTCCCGTGCAGCAGCATCCACAGCGAGGGGCGAGTGGGTGAGC 3255
Db 2800 CCACCCCCACCTGACGCTCCCGTGCAGCAGCATCCACAGCGAGGGGCGAGTGGGTGAGC 2859
QY 3256 CTGCGGGCCCTTGGGACACCATCAACGTCCACCTCCGGGCTGGGTATCATATCCCCCTG 3315
Db 2860 CTGCGGGCCCTTGGGACACCATCAACGTCCACCTCCGGGCTGGGTATCATATCCCCCTG 2919
QY 3316 CAGGGCCCTTGGGCTCACAAACAGAGTCCCGCAGCAGCCCATGCGCCCTGCTGTCGCC 3375
Db 2920 CAGGGCCCTTGGGCTCACAAACAGAGTCCCGCAGCAGCCCATGCGCCCTGCTGTCGCC 2979
QY 3376 CTAAACAAAGGTGGAGAGCGCCGAGGGAGCTGTTCTGGACGATGGAGAGCCTGGAA 3435
Db 2980 CTGACCAAGGTGGAGAGCGCCGAGGGAGCTGTTCTGGACGATGGAGAGCCTGGAA 3039
QY 3436 GTGCTGGAGCGAGGGGCTTACACAGGTATCTTCTTGGCCAGGAATAACACGATCGTG 3495
Db 3040 GTGCTGGAGCGAGGGGCTTACACAGGTATCTTCTTGGCCAGGAATAACACGATCGTG 3099
QY 3496 AATGAGCTGTGATGTTGACCAAGTGCAGGAGCTGGCCCTGCAGCTGCAGAGGTGACTCTC 3555
Db 3100 AATGAGCTGTGATGTTGACCAAGTGCAGGAGCTGGCCCTGCAGCTGCAGAGGTGACTCTC 3159
QY 3556 CTGGCGTGGGCGACCGCGCCCGCAGCAGGTCTCTCCAAAGGTGCTCTGCTTCCAACTTC 3615
Db 3160 CTGGCGTGGGCGACCGCGCCCGCAGCAGGTCTCTCCAAAGGTGCTCTGCTTCCAACTTC 3219
QY 3616 ACCTACAGCCCGACACCAAGTCTGACATCTGTGCTGCTGATGGAGAGCAG 3675
Db 3220 ACCTACAGCCCGACACCAAGTCTGACATCTGACATCTGTGCTGCTGATGGAGAGCAG 3279
QY 3676 TTCTCTGTCAGTGGTGT 3694
Db 3280 TTCTCTGTCAGTGGTGT 3298

RESULT 4
LOCUS CQ723359
DEFINITION Sequence 9293 from Patent WO02068579.
ACCESSION CQ723359
VERSION CQ723359.1 GI:42284216
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kites, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses

CQ723359 3625 bp DNA linear PAT 03-FEB-2004
Sequence 9293 from Patent WO02068579.

QY 3016 CTCCTCTCGAGTTCCCAAGACTTAGCACTGGAAGTGTGGACCAACAGCTCTCTGTGG 3075
Db CTCCTCTCGAGTTCCCAAGACTTAGCACTGGAAGTGTGGACCAACAGCTCTCTGTGG 2456
QY 3076 GGGAGGCGCTGCTCATACCCAGTGTCTCAGGCGGGAAGCCGAAAGTGAAGTGAAGTGA 3135
Db GGGAGGCGCTGCTCATACCCAGTGTCTCAGGCGGGAAGCCGAAAGTGAAGTGAAGTGA 2516
QY 3136 TTCCCTCTCGGACATGTTAGCACTGCGAGCGGTGCAATAGAGCCCTTGGAGCCCTC 3195
Db TTCCCTCTCGGACATGTTAGCACTGCGAGCGGTGCAATAGAGCCCTTGGAGCCCTC 2576
QY 3196 CCACCCCACTGAGCTCCCGTGGAGCGAGCCATCCACGAGGAGGGGAGTGGGTGAGC 3255
Db CCACCCCACTGAGCTCCCGTGGAGCGAGCCATCCACGAGGAGGGGAGTGGGTGAGC 2636
QY 3256 CTGCGGCGCCCTTGAGACCAATCAAGCTCAGCTCCGGGCTGGGTACATCATCTCCCTG 3315
Db CTGCGGCGCCCTTGAGACCAATCAAGCTCAGCTCCGGGCTGGGTACATCATCTCCCTG 2696
QY 3316 CAGGCGCTGGCTCAACACACAGAGTCCCGCAGCAGCCATGCGCTGTGGCC 3375
Db CAGGCGCTGGCTCAACACACAGAGTCCCGCAGCAGCCATGCGCTGTGGCC 2756
QY 3376 CTAACCAAGGCTGGAGGCGCCGAGGGAGCTGTTCTCGGACGATGGAGAGCCCTGGAA 3435
Db CTAACCAAGGCTGGAGGCGCCGAGGGAGCTGTTCTCGGACGATGGAGAGCCCTGGAA 2816
QY 3436 GTGCTGGAGGAGGCGCTACACAGGTATCTTCTGCGCAGGATTAACAGATCGTG 3495
Db GTGCTGGAGGAGGCGCTACACAGGTATCTTCTGCGCAGGATTAACAGATCGTG 2876
QY 3496 AATGAGCTGTAGTGTGACCAAGTGAAGGAGCTGGCTGCGCTGCGAGAGTGAAGTGA 3555
Db AATGAGCTGTAGTGTGACCAAGTGAAGGAGCTGGCTGCGCTGCGAGAGTGAAGTGA 2936
QY 3556 CTGGGCGTGGCCACGCGCGCCGAGAGGTCCTTCCAAAGTGTCTGCTTCCAACTTC 3615
Db CTGGGCGTGGCCACGCGCGCCGAGAGGTCCTTCCAAAGTGTCTGCTTCCAACTTC 2996
QY 3616 ACCTAGAGCGCGACACCAAGTGTCTGAGACATCTGTCTCGTGTGATGGAGAGCAG 3675
Db ACCTAGAGCGCGACACCAAGTGTCTGAGACATCTGTCTCGTGTGATGGAGAGCAG 3056
QY 3676 TTCTCTGTCAGTGGTGT 3694
Db TTCTCTGTCAGTGGTGT 3075

RESULT 5
HUMGAAA
LOCUS Human acid alpha-glucosidase (GAA) mRNA, complete cds. PRI 08-NOV-1994
DEFINITION
ACCESSION M34424
VERSION M34424.1
KEYWORDS acid alpha-glucosidase; acid maltase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3441)
AUTHORS Martinik, F., Mehler, M., Tzall, S., Meredith, G. and Hirschhorn, R.
TITLE Sequence of the cDNA and 5'-flanking region for human acid alpha-glucosidase, detection of an intron in the 5' untranslated leader sequence, definition of 18-bp polymorphisms, and differences with previous cDNA and amino acid sequences
JOURNAL DNA 9, 85-94 (1990)
COMMENT Original source text: Human, cDNA to mRNA.
FEATURES
Location/Qualifiers
1..3441
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

/map="17q23"
1..3441
/gene="GAA"
33..2891
/gene="GAA"
/note="acid alpha-glucosidase"
/codon_start=1
/protein_id="AAA52506.1"
/db_xref="GI:182908"
/db_xref="GDB:G00-119-965"
/translation="MGVRRPCSHLLAVCALVSLATALLGHILLHDFLLVPRLSG
SSVLETHPAHQGASRGRDRAQHPRAVPTQCDVPNSRGTADPDKAITPTEQ
CEARCCYIPAKQGLQGMQPMCFPPSPYKLENSSEMSEMYTALTTRFTTFE
PKDILTRLDVMMETENRHTIKDPAANRVEVPLETPHVSRAPSPLYSRSEPEF
QVIVROLDGRVLLNTVPLFFADQLOLSTSLPSOITGLAEHLSPLMLSTWTRI
TLNWRDLAPTGANLYGSHPFYLALEDGSGAHGVFLNANMDVVLQPSPALSWSTG
GILDYIFLGPBPKSVQOQYLDVGVYPMPPYWGVLGFLCRNGYSSTAITRQVVENMT
RAHFFYDVQWDLDDMSRDFTFNKGRDFPAMVQELHQGGRRYMTIIVPAISGG
PAGSYRPFDEGLRRGVFTNETGQPLIGKWPSTAFDPFTNPTALWMDVABFHD
QVPFGMDIMNENPSFIRGSDGCPNNELENPPYVGVGTIQAATICASHQFLS
THYNLNLGLTEAASHRALVKARTRPPVLSRSTFAGHGRYAGHWTGDVMSWEOL
ASSVPEILQFNLLGVPLVADVCPLGNTSELCVRWTOLGAFVPMENHNSLLSLPQ
EPYSFSEPAQAKRKALILRLYLLPHLYTLFQHAHAGETVARPLPLFFPKDSSWTYV
DHQLLWGBALLITPVLQAGKAEVTGYFPLGTWYDLQTVFIEALGSLVPLPPAPRPPI
HSEQWVTLPLDPTINVHLRAGYIIPLOGLTITRESRQPMALVALITKGGEARGE
LFWDDGESLEVLERGAYTVIFLARNTIIVNELVRVTSEGAGLOLQKVTVLGATAPQ
QVLSNGVPVSNFTYSPDTKVLIDICVSLLMGBOFLVSWC"

ORIGIN

Query Match 71.3%; Score 2639.8; DB 9; Length 3441;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2647; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1036 GAAACCGGTGCACACCCCGCGCTCCACAGAGCAGTCCACACAGTGGAGCTGCCCC 1095
Db 231 GATGCCAGGCACACCCCGCGCTCCACAGAGCAGTCCACACAGTGGAGCTGCCCC 290
QY 1096 AACAGCCGCTTCGATTGGCGCCCTGACAAAGGCCATCACCAGAGAACAGTGCAGAGCCCGC 1155
Db 291 AACAGCCGCTTCGATTGGCGCCCTGACAAAGGCCATCACCAGAGAACAGTGCAGAGCCCGC 350
QY 1156 GCGTCTGCTACATCCCTGCAAGCAGGGGTGAGGAGGCCAGATGGGGCAGCCCTGG 1215
Db 351 GCGTCTGCTACATCCCTGCAAGCAGGGGTGAGGAGGCCAGATGGGGCAGCCCTGG 410
QY 1216 TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGAGAGAACCTGAGCTCTCTGAAATG 1275
Db 411 TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGAGAGAACCTGAGCTCTCTGAAATG 470
QY 1276 GGTACACGGCCACCTGACCCCGTACACCCCGCTTCTTCTTCCCAAGGACATCTGACC 1335
Db 471 GGTACACGGCCACCTGACCCCGTACACCCCGCTTCTTCTTCCCAAGGACATCTGACC 530
QY 1336 CTGCGGCTGAGAGTATGATGAGAGCTGAGAACCCCTCCACTTCACGATCAAGATCCA 1395
Db 531 CTGCGGCTGAGAGTATGATGAGAGCTGAGAACCCCTCCACTTCACGATCAAGATCCA 590
QY 1396 GCTACAGCGCTACGAGGTGCCCTTGGAGACCCCGCTGTCCACAGCGGGSCACCGTCC 1455
Db 591 GCTACAGCGCTACGAGGTGCCCTTGGAGACCCCGCTGTCCACAGCGGGSCACCGTCC 650
QY 1456 CCACCTCTACAGCTGGAGTTCCTCCGAGGAGCCCTTCCGGGGTGATCGTGACCGGAGCTG 1515
Db 651 CCACCTCTACAGCTGGAGTTCCTCCGAGGAGCCCTTCCGGGGTGATCGTGACCGGAGCTG 710
QY 1516 GACGCGCGGTGCTGCTGAAACACGACGCTGGGGCCCTTCTTTTGGGACAGTTCCTT 1575
Db 711 GACGCGCGGTGCTGCTGAAACACGACGCTGGGGCCCTTCTTTTGGGACAGTTCCTT 770
QY 1576 CAGCTGTCCACCTCGCTCGCTCGAGTATATACAGGCTGCGCGGAGCACCTCAGTCC 1635
Db 771 CAGCTGTCCACCTCGCTCGCTCGAGTATATACAGGCTGCGCGGAGCACCTCAGTCC 830

Qy	1636	CTGATGCTCAGCACACGCTGGACAGGATCACCTGTGGAAACGGGACCTTGCGCCACG	1695
Db	831	CTGATGCTCAGCACACGCTGGACAGGATCACCTGTGGAAACGGGACCTTGCGCCACG	890
Qy	1696	CCGGTGCGAACCTCTACGGGTCTCACCCCTTTCTACCTGGCGCTGGAGGACGGCGGGTCG	1755
Db	891	CCGGTGCGAACCTCTACGGGTCTCACCCCTTTCTACCTGGCGCTGGAGGACGGCGGGTCG	950
Qy	1756	GCACACGGGGTGTTCCTGCTAAACAGACAAATGCCATGGATGTGGTCTCTGACAGCCGAGCCCT	1815
Db	951	GCACACGGGGTGTTCCTGCTAAACAGACAAATGCCATGGATGTGGTCTCTGACAGCCGAGCCCT	1010
Qy	1816	GCCTTAGCTGGAGGTCCGACAGGTGGGATCCTCGATGTCTACATCTTCCTGGGCCCAGAG	1875
Db	1011	GCCTTAGCTGGAGGTCCGACAGGTGGGATCCTCGATGTCTACATCTTCCTGGGCCCAGAG	1070
Qy	1876	CCCNAGAGCGTGTGACAGAGTACCTGGACGTTGTGGGATACCGGTTTCATGCGCCCATAC	1935
Db	1071	CCCNAGAGCGTGTGACAGAGTACCTGGACGTTGTGGGATACCGGTTTCATGCGCCCATAC	1130
Qy	1936	TGGGGCTTGGCTTCACACCTGTGCGGTGGGGTACTCTCTCCACCGCTATCATCCCGCCAG	1995
Db	1131	TGGGGCTTGGCTTCACACCTGTGCGGTGGGGTACTCTCTCCACCGCTATCATCCCGCCAG	1190
Qy	1996	GTGTTGGAGAACATGACAGGGGCCACTTCCCCTTGGACGTCAAATGGAAACACTGGAC	2055
Db	1191	GTGTTGGAGAACATGACAGGGGCCACTTCCCCTTGGACGTCAAATGGAAACACTGGAC	1250
Qy	2056	TACATGACTTCCGGAGGGACTTCAGGTTCAAACAGGATGGCTTCCGGGACTTCCCGGCC	2115
Db	1251	TACATGACTTCCGGAGGGACTTCAGGTTCAAACAGGATGGCTTCCGGGACTTCCCGGCC	1310
Qy	2116	ATGTTGACGAGCTGACACAGGGCGGCGCGGCTACATGATGATCGTGGATCCTGCATC	2175
Db	1311	ATGTTGACGAGCTGACACAGGGCGGCGCGGCTACATGATGATCGTGGATCCTGCATC	1370
Qy	2176	AGCAGCTCGGGCCTTGCCTGGAGACTACAGGCCCTTACGACGAGGCTTTCGCGAGGGGGTT	2235
Db	1371	AGCAGCTCGGGCCTTGCCTGGAGACTACAGGCCCTTACGACGAGGCTTTCGCGAGGGGGTT	1430
Qy	2236	TTCAATACCAACGAGACCGGCCAGCCGCTGATTTGGGAGGTATGGCCCGGGTCCACTGCC	2295
Db	1431	TTCAATACCAACGAGACCGGCCAGCCGCTGATTTGGGAGGTATGGCCCGGGTCCACTGCC	1490
Qy	2296	TTCCCCGACTTCACCAACCCCAACAGCCCTGGCTGGTGGAGGACATGGTGGCTGAGTTC	2355
Db	1491	TTCCCCGACTTCACCAACCCCAACAGCCCTGGCTGGTGGAGGACATGGTGGCTGAGTTC	1550
Qy	2356	CATGACAGGTCGCCCTTCGACGGCTTGTGATTGACATGAACGAGCCTTCCAACTTCAATC	2415
Db	1551	CATGACAGGTCGCCCTTCGACGGCTTGTGATTGACATGAACGAGCCTTCCAACTTCAATC	1610
Qy	2416	AGAGGCTCTGAGGACGGCTGCCCCAAACAATGAGCTGGAGAAACCCACCTACGTGCCTGGG	2475
Db	1611	AGAGGCTCTGAGGACGGCTGCCCCAAACAATGAGCTGGAGAAACCCACCTACGTGCCTGGG	1670
Qy	2476	GTGTTTGGGGGACCCCTCAGGGGGCCACCATCTGTGCCTCCAGACCAACAGTTTCTCTCC	2535
Db	1671	GTGTTTGGGGGACCCCTCAGGGGGCCACCATCTGTGCCTCCAGACCAACAGTTTCTCTCC	1730
Qy	2536	ACACACTAACCTTGCAACACTTACGGCCTGACCGAAGCCATCGCCTCCCAAGGGGG	2595
Db	1731	ACACACTAACCTTGCAACACTTACGGCCTGACCGAAGCCATCGCCTCCCAAGGGGG	1790
Qy	2596	CTGGTGAAGGCTCGGGGGAACGCCCAATTTGTGATCTCCCGCTCGACCTTGTGCGCCAC	2655
Db	1791	CTGGTGAAGGCTCGGGGGAACGCCCAATTTGTGATCTCCCGCTCGACCTTGTGCGCCAC	1850
Qy	2656	GGCCGATACCGCGGCCACTTGGACGGGGGACGTGTGGAGCTCCTGGGAGAGACTCGCTCC	2715
Db	1851	GGCCGATACCGCGGCCACTTGGACGGGGGACGTGTGGAGCTCCTGGGAGAGACTCGCTCC	1910
Qy	2716	TCGTTGCCAGAAATTCCTGCATTTAACTCTCGGGGTGCCTCTGTGTCGGGGCCGACGT	2775

RESULT 6
BC040431
LOCUS
DEFINITION

IMAGE:5739991), complete cds.	/codon_start=1
BC040431	/product="GAA protein"
VERSION	/protein_id="AAH40431.1"
KEYWORDS	/db_xref="GI:26251857"
SOURCE	/db_xref="GI:26251857"
ORGANISM	/db_xref="LocusID:2548"
Homo sapiens (human)	/db_xref="MIM:606800"
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 3687)	
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,	
Klauser,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,	
Altschul,S.F., Zebberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,	
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,E.,	
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,	
Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,	
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,	
Carninci,P., Prange,C., Roha,S.S., Loquellano,N.A., Peters,G.J.,	
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,	
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.W.,	
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,	
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,	
Fahy,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,	
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,	
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,	
Dickson,M.C., Rodriguez,A.C., Grinnwood,J., Schmutz,J., Myers,R.M.,	
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,	
Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.	
Generation and initial analysis of more than 15,000 full-length	
human and mouse cDNA sequences	
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
22398257	
12477932	
2 (bases 1 to 3687)	
Strausberg,R.	
Direct Submission	
Submitted (29-NOV-2002) National Institutes of Health, Mammalian	
Gene Collection (MGC), Cancer Genomics Office, National Cancer	
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,	
USA	
NIH-MGC Project URL: http://mgc.nci.nih.gov	
Contact: MGC help desk	
Email: cgabs-remail.nih.gov	
Tissue Procurement: ATCC	
cDNA Library Preparation: Life Technologies, Inc.	
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
DNA Sequencing by: Institute for Systems Biology	
http://www.systemsbio.org	
contact: amadnan@systemsbiology.org	
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha	
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting	
Clone distribution: MGC clone distribution information can be found	
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	
Series: IRAC Plate: 84 Row: j Column: 8	
This clone was selected for full length sequencing because it	
passed the following selection criteria: matched mRNA gi: 11496988.	
Location/Qualifiers	
1. .3687	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="MGC:48746 IMAGE:5739991"	
/tissue_type="Duodenum, adenocarcinoma"	
/clone_lib="NIH MGC_88"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .3687	
/gene="GAA"	
/note="synonym: LVAG"	
/db_xref="LocusID:2548"	
/db_xref="MIM:606800"	
145. .3003	
/gene="GAA"	
CDS	
1036 GAAACGGTGCACACCCCGCGCTCCAGAGAGTGCACACAGTGCAGCTCCCGCC 1095	Query Match 71.2%; Score 2635; DB 9; Length 3687;
	Best Local Similarity 99.4%; Pred. No. 0;
343 GATGCCAGGCACACCCCGCGCTCCAGAGAGTGCACACAGTGCAGCTCCCGCC 402	Matches 2644; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
1096 AACACCGCTTCGATTGCGCCCTGCACAAAGGCATCATCCAGAAACAGTGCAGGCCGCG 1155	
403 AACACCGCTTCGATTGCGCCCTGCACAAAGGCATCATCCAGAAACAGTGCAGGCCGCG 462	
1156 GGCTCTCTACATCCCTGCACAAAGCAGGGGTGCAGGAGGCCAGATGGGGCAGCCCTGG 1215	
463 GGCTCTCTACATCCCTGCACAAAGCAGGGGTGCAGGAGGCCAGATGGGGCAGCCCTGG 522	
1216 TGCTTTCTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTGAAATG 1275	
523 TGCTTTCTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTGAAATG 582	
1276 GGCTACAGCGCCACCTGACCCGCTACACCCCGCTTCTTCCCAAGGACATCCTGACC 1335	
583 GGCTACAGCGCCACCTGACCCGCTACACCCCGCTTCTTCCCAAGGACATCCTGACC 642	
1336 CTGGCGCTGGAGCTGATGATGAGACTGAGAACCCGCTCCACTTCACGATCAAGATCCA 1395	
643 CTGGCGCTGGAGCTGATGATGAGACTGAGAACCCGCTCCACTTCACGATCAAGATCCA 702	
1396 GCTACAGCGCTGACAGTGCCTTGGAGACCCCGCTGTCACAGCGCGGACCGCTCC 1455	
703 GCTACAGCGCTGACAGTGCCTTGGAGACCCCGCTGTCACAGCGCGGACCGCTCC 762	
1456 CCACCTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGGTGATCGTCCACCGGAGCTG 1515	
763 CCACCTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGGTGATCGTCCACCGGAGCTG 822	
1516 GACGGCGCGTGTCTGCTGAAACACAGAGGTGGCGCCCTGTCTTTTGGCGACCGATTCCTT 1575	
ORIGIN	
misc_feature	
385. .546	
/gene="GAA"	
/note="PD; Region: P or trefoil or TFF domain"	
/db_xref="CDD:smart00018"	
724. .2895	
/gene="GAA"	
/note="Glyco_hydro_31; Region: Glycosyl hydrolases family	
31. Glycosyl hydrolases are key enzymes of carbohydrate	
metabolism. Family 31 comprises of enzymes that are, or	
similar to, alpha-galactosidases"	
/db_xref="CDD:pfam01055"	
misc_feature	
724. .2895	
/gene="GAA"	
/note="Glyco_hydro_31; Region: Glycosyl hydrolases family	
31. Glycosyl hydrolases are key enzymes of carbohydrate	
metabolism. Family 31 comprises of enzymes that are, or	
similar to, alpha-galactosidases"	
/db_xref="CDD:pfam01055"	
FEATURES	
source	
1. .3687	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone="MGC:48746 IMAGE:5739991"	
/tissue_type="Duodenum, adenocarcinoma"	
/clone_lib="NIH MGC_88"	
/lab_host="DH10B"	
/note="Vector: pCMV-SPORT6"	
1. .3687	
/gene="GAA"	
/note="synonym: LVAG"	
/db_xref="LocusID:2548"	
/db_xref="MIM:606800"	
145. .3003	
/gene="GAA"	
CDS	
1036 GAAACGGTGCACACCCCGCGCTCCAGAGAGTGCACACAGTGCAGCTCCCGCC 1095	
343 GATGCCAGGCACACCCCGCGCTCCAGAGAGTGCACACAGTGCAGCTCCCGCC 402	
1096 AACACCGCTTCGATTGCGCCCTGCACAAAGGCATCATCCAGAAACAGTGCAGGCCGCG 1155	
403 AACACCGCTTCGATTGCGCCCTGCACAAAGGCATCATCCAGAAACAGTGCAGGCCGCG 462	
1156 GGCTCTCTACATCCCTGCACAAAGCAGGGGTGCAGGAGGCCAGATGGGGCAGCCCTGG 1215	
463 GGCTCTCTACATCCCTGCACAAAGCAGGGGTGCAGGAGGCCAGATGGGGCAGCCCTGG 522	
1216 TGCTTTCTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTGAAATG 1275	
523 TGCTTTCTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTGAAATG 582	
1276 GGCTACAGCGCCACCTGACCCGCTACACCCCGCTTCTTCCCAAGGACATCCTGACC 1335	
583 GGCTACAGCGCCACCTGACCCGCTACACCCCGCTTCTTCCCAAGGACATCCTGACC 642	
1336 CTGGCGCTGGAGCTGATGATGAGACTGAGAACCCGCTCCACTTCACGATCAAGATCCA 1395	
643 CTGGCGCTGGAGCTGATGATGAGACTGAGAACCCGCTCCACTTCACGATCAAGATCCA 702	
1396 GCTACAGCGCTGACAGTGCCTTGGAGACCCCGCTGTCACAGCGCGGACCGCTCC 1455	
703 GCTACAGCGCTGACAGTGCCTTGGAGACCCCGCTGTCACAGCGCGGACCGCTCC 762	
1456 CCACCTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGGTGATCGTCCACCGGAGCTG 1515	
763 CCACCTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGGTGATCGTCCACCGGAGCTG 822	
1516 GACGGCGCGTGTCTGCTGAAACACAGAGGTGGCGCCCTGTCTTTTGGCGACCGATTCCTT 1575	

Db 823 |||||GACGGCCGGTGTGTGAACGACGAGTGGCGCCCTGTCTTTTCGGACAGTTCTCTT 882
QY 1576 CAGCTGTCCACCTTCGCTGCCCTCGCAGTATATCACAGGCTTCGCGAGCACCTCAGTCCC 1635
Db 883 CAGCTGTCCACCTTCGCTGCCCTCGCAGTATATCACAGGCTTCGCGAGCACCTCAGTCCC 942
QY 1636 CTGATGCTCAGCACCAAGCTGGACAGGATCACCTGTGGAAACGGGACCTTGGCCACG 1695
Db 943 CTGATGCTCAGCACCAAGCTGGACAGGATCACCTGTGGAAACGGGACCTTGGCCACG 1002
QY 1696 CCGGTGCGAACTCTACGGGTCTACCTTTCTACTTGGCGCTGAGACGGCGGGTCG 1755
Db 1003 CCGGTGCGAACTCTACGGGTCTACCTTTCTACTTGGCGCTGAGACGGCGGGTCG 1062
QY 1756 GCACAGGGGTGTCTCTGTAAACAGCAATGCCATGATGTGTCTGTGAGCGGAGCCCT 1815
Db 1063 GCACAGGGGTGTCTCTGTAAACAGCAATGCCATGATGTGTCTGTGAGCGGAGCCCT 1122
QY 1816 GCCCTTAGCTGGAGGTGACAGGTGGGATCTTGATGTCTACATCTTCTGGGCCACAG 1875
Db 1123 GCCCTTAGCTGGAGGTGACAGGTGGGATCTTGATGTCTACATCTTCTGGGCCACAG 1182
QY 1876 CCCAAGAGCTGTGACAGGATACCTGGAGTGTGTGGATATACCGTTCATGTGCGCCATAC 1935
Db 1183 CCCAAGAGCTGTGACAGGATACCTGGAGTGTGTGGATATACCGTTCATGTGCGCCATAC 1242
QY 1936 TGGGGCTCGGGCTTCCACCTGTGCGCTGGGGTACTCTCCACCGCTATCACCGGCCAG 1995
Db 1243 TGGGGCTCGGGCTTCCACCTGTGCGCTGGGGTACTCTCTCCACCGCTATCACCGGCCAG 1302
QY 1996 GTGGTGGAAACATGACAGGGGCCCACTTCCCTCGAGCTCCAAATGGAAACGACCTGGAC 2055
Db 1303 GTGGTGGAAACATGACAGGGGCCCACTTCCCTCGAGCTCCAAATGGAAACGACCTGGAC 1362
QY 2056 TACATGGAATCCGGAGGACTTACAGTTCAACAGGATGGCTTCGGGACTTCCGGGCC 2115
Db 1363 TACATGGAATCCGGAGGACTTACAGTTCAACAGGATGGCTTCGGGACTTCCGGGCC 1422
QY 2116 ATGGTGCAGAGCTGCACAGGGCGCGCGCTACATGATGATCGTGATCCTGCCATC 2175
Db 1423 ATGGTGCAGAGCTGCACAGGGCGCGCGCTACATGATGATCGTGATCCTGCCATC 1482
QY 2176 AGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTACAGCAGGGTCTGGGAGGGGGTT 2235
Db 1483 AGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTACAGCAGGGTCTGGGAGGGGGTT 1542
QY 2236 TTTCATCAACAGAGACCGGCCAGCGCTGATTTGGGAAGGATGGCCCGGGTCCATGCC 2295
Db 1543 TTTCATCAACAGAGACCGGCCAGCGCTGATTTGGGAAGGATGGCCCGGGTCCATGCC 1602
QY 2296 TTTCCTCGACTTCACCAACCCACAGCCCTGGCTGTGGAGGACATGGTGGTGAAGTTC 2355
Db 1603 TTTCCTCGACTTCACCAACCCACAGCCCTGGCTGTGGAGGACATGGTGGTGAAGTTC 1662
QY 2356 CATGACAGGTGCTTCGACGGCTTGTGATGATGATGAACAGAGCTTCCAACTTCATC 2415
Db 1663 CATGACAGGTGCTTCGACGGCTTGTGATGATGATGAACAGAGCTTCCAACTTCATC 1722
QY 2416 AGAGGTCTTAGAGACGGCTGCCCAACATAGTGGAGAACCCATCCTAGTGGCTGGG 2475
Db 1723 AGGGGTCTTAGAGACGGCTGCCCAACATAGTGGAGAACCCATCCTAGTGGCTGGG 1782
QY 2476 GTGGTTGGGGGAGCCCTCCAGGGGCCCATCTGTGCTCCAGCCACCAAGTTCTCTCC 2535
Db 1783 GTGGTTGGGGGAGCCCTCCAGGGGCCCATCTGTGCTCCAGCCACCAAGTTCTCTCC 1842
QY 2536 ACACACTACAACTGCACAACTCTACGGCTGACCGAAGCCATCGCTCCCAAGGGCG 2595
Db 1843 ACACACTACAACTGCACAACTCTACGGCTGACCGAAGCCATCGCTCCCAAGGGCG 1902
QY 2596 CTGGTGAAGGCTCGGGGAGACGCCCAATTTGATCTCCCGTTCGACCTTTGTGGCCAC 2655
|||||

Db 1903 CTGGTGAAGGCTCGGGGAGACGCCCAATTTGTGATCTCCCGCTCGACCTTTGTGTCGCAC 1962
QY 2656 GGGCCATACCGCGGCCACTGGACGGGGACGTGTGTGAGCTCTCTGGAGCAGCTCGCTCC 2715
Db 1963 GGGCCATACCGCGGCCACTGGACGGGGACGTGTGTGAGCTCTCTGGAGCAGCTCGCTCC 2022
QY 2716 TCCGTGCCAGAAATCCTGCAAGTTTAACTGTGTGGGGTGCTCTGTGTGGGGCCGACGTC 2775
Db 2023 TCCGTGCCAGAAATCCTGCAAGTTTAACTGTGTGGGGTGCTCTGTGTGGGGCCGACGTC 2082
QY 2776 TCGGCTTCTCGGGCAACACCTCAGAGGAGCTGTGTGTGCGTGGACCCAGCTGGGGCC 2835
Db 2083 TCGGCTTCTCGGGCAACACCTCAGAGGAGCTGTGTGTGCGTGGACCCAGCTGGGGCC 2142
QY 2836 TTCTACCCCTTCATGCGGAACCAACAGCTGTCTCAGTCTGCCCCAGAGCCGTACAGC 2895
Db 2143 TTCTACCCCTTCATGCGGAACCAACAGCTGTCTCAGTCTGCCCCAGAGCCGTACAGC 2202
QY 2896 TTCAGGAGCCGGCCAGCAGGCCATGAGGAAGGGCCCTCACCTTGGCTTACGCACTCTC 2955
Db 2203 TTCAGGAGCCGGCCAGCAGGCCATGAGGAAGGGCCCTCACCTTGGCTTACGCACTCTC 2262
QY 2956 CCCCACCTCTACACACTGTTCACACAGCCCACTGTGCGGGGAGACCGTGGCCCGGCC 3015
Db 2263 CCCCACCTCTACACACTGTTCACACAGGCCCACTGTGCGGGGAGACCGTGGCCCGGCC 2322
QY 3016 CTCTTCTGGAGTTCCTCCAAAGGACTTAGCACCTTGGACTGTGTGGACCCAGCTCTGTGG 3075
Db 2323 CTCTTCTGGAGTTCCTCCAAAGGACTTAGCACCTTGGACTGTGTGGACCCAGCTCTGTGG 2382
QY 3076 GGGAGGGCCCTCTCATCACTGTCTCAGGCCGGGAAGGCCGAAAGTGAATGCTGTCTAC 3135
Db 2383 GGGAGGGCCCTCTCATCACTGTCTCAGGCCGGGAAGGCCGAAAGTGAATGCTGTCTAC 2442
QY 3136 TTCTCTTGGGACATGTATGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCCTC 3195
Db 2443 TTCTCTTGGGACATGTATGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCCTC 2502
QY 3196 CCACCCCACTGACGTCTCCCGTGAAGCCAGCCATCCACAGAGGGGAGTGGGTGAGC 3255
Db 2503 CCACCCCACTGACGTCTCCCGTGAAGCCAGCCATCCACAGAGGGGAGTGGGTGAGC 2562
QY 3256 CTGGCGGGCCCTCTGGACACATCAACGTCTCACTCCGGGCTGGGTATCATCATCCCCCTG 3315
Db 2563 CTGGCGGGCCCTCTGGACACATCAACGTCTCACTCCGGGCTGGGTATCATCATCCCCCTG 2622
QY 3316 CAGGGCCCTGGCTCACAAACACAGAGTCCCGCAGAGCCCATGSCCTGGCTGTGGCC 3375
Db 2623 CAGGGCCCTGGCTCACAAACACAGAGTCCCGCAGAGCCCATGSCCTGGCTGTGGCC 2682
QY 3376 CTAAACAAAGGGTGGAGAGGCCCGAGGGAGCTGTCTTGGGACGATGGAGAGCCCTGGAA 3435
Db 2683 CTGACCAAGGGTGGGAGGGCCCGAGGGAGCTTTTCTGGGACGATGGAGAGCCCTGGAA 2742
QY 3436 GTGCTGGAGCGAGGGCCCTTACACAGGTCTCTCTTCTGGCCAGGAATTAACAGATCGTG 3495
Db 2743 GTGCTGGAGCGAGGGCCCTTACACAGGTCTCTCTTCTGGCCAGGAATTAACAGATCGTG 2802
QY 3496 AATGAGCTGGTACGTGTGACAGTGGAGGCTGGCCCTGACCTGCAGAGGAGTGAAGTTC 3555
Db 2803 AATGAGCTGGTACGTGTGACAGTGGAGGCTGGCCCTGACCTGCAGAGGAGTGAAGTTC 2862
QY 3556 CTGGGCTGGCCACCGCGGCCCGCAGAGGTCTCTCCACGGTGTCTCTCTTCCAACTTC 3615
Db 2863 CTGGGCTGGCCACCGCGGCCCGCAGAGGTCTCTCTCCACGGTGTCTCTCTTCCAACTTC 2922
QY 3616 ACCTACAGCCCGACACCAAGGTCTTGGACATCTGTGTCTCGCTGTGTGATGGAGAGCAG 3675
Db 2923 ACCTACAGCCCGACACCAAGGTCTTGGACATCTGTGTCTCGCTGTGTGATGGAGAGCAG 2982
QY 3676 TTTCTCGTCAGCTGGTGT 3694
Db 2983 TTTCTCGTCAGCTGGTGT 3001


```
RESULT 7
LOCUS BV177050/c 3626 bp DNA linear STS 10-JUN-2004
DEFINITION sqm92407 Human DNA (Sequenom) Homo sapiens STS genomic, sequence
tagged site.
ACCESSION BV177050
VERSION BV177050
KEYWORDS STS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3626)
AUTHORS Nelson,R.M., Marnellos,G., Kammerer,S., Hoyal,C.R., Shi,M.M.,
Cantor,C.R. and Braun,A.
TITLE Large-Scale Validation of Single Nucleotide Polymorphisms in Gene
Regions
JOURNAL Genome Res. (2004) In press
COMMENT Contact: Andreas Braun
Pharmaceuticals division
Sequenom, Inc.
3595 John Hopkins Court, San Diego, CA 92121, USA
Tel: 18582029018
Fax: 18582029020
Email: abraun@sequenom.com
Primer A: No primer sequence submitted
Primer B: No primer sequence submitted
STS size: 3626.

FEATURES
source
location/Qualifiers
1..3626
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="Human DNA (Sequenom)"
<1..>3626

ORIGIN
Query Match 69.8%; Score 2585.8; DB 11; Length 3626;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2640; Conservative 0; Mismatches 17; Indels 4; Gaps 4;

QY 1036 GAAACCGGTGCACACCCCGCGCTCCAGAGCAGTGCCACACAGTGCGACGTCCCGCCCC 1095
DB |||
DB 3209 GATGCCAGGCACACCCCGCGCTCCAGAGCAGTGCCACACAGTGCGACGTCCCGCCCC 3150
QY 1096 AACAGCCGTTGATTTGCGCCCTGACAAAGCCATCACCCAGGAAACAGTGCAGGCCCGC 1155
DB |||
DB 3149 AACAGCCGTTGATTTGCGCCCTGACAAAGGCCATCACCCAGGAAACAGTGCAGGCCCGC 3090
QY 1156 GCGTGTGTACATCCCTGCAAGCAGGCGCTGCAGGAGCCAGATGGGCGAGCCCTGG 1215
DB |||
DB 3089 GCGTGTGTACATCCCTGCAAGCAGGCGCTGCAGGAGCCAGATGGGCGAGCCCTGG 3030
QY 1216 TGGTTCTTCCACCCAGCTACAGCTGGAGAACTGAGCTCTCTGAAATG 1275
DB |||
DB 3029 TGGTTCTTCCACCCAGCTACAGCTGGAGAACTGAGCTCTCTGAAATG 2970
QY 1276 GGTACACGGCCACCTGTACCCCGTACACCCCGACCTTCTTCCCGAAGACATCTGTACC 1335
DB |||
DB 2969 GGTACACGGCCACCTGTACCCCGTACACCCCGACCTTCTTCCCGAAGACATCTGTACC 2910
QY 1336 CTGCGGCTGAGTGTATGGAGCTGAGACCGGCTCCACTTCAGATCAAGATCCA 1395
DB |||
DB 2909 CTGCGGCTGAGTGTATGGAGCTGAGACCGGCTCCACTTCAGATCAAGATCCA 2850
QY 1396 GGTAAACAGGCGTACAGGTTGCCCTTTGGAGACCCCGCGTGTCCACAGCCGGGACCGTCC 1455
DB |||
DB 2849 GCTAACAGGCGTACAGGTTGCCCTTTGGAGACCCCGCGTGTCCACAGCCGGGACCGTCC 2790
QY 1456 CCACCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGCGGGTGATCGTGCAACGGCAGCTG 1515

2789 CCACCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGCGGGTGATCGTGCAACCGGCAGCTG 2730
1516 GACGCGCGGTCTCTCTGAAACACGACGCGTGGCGGCCCTTCTTTTGGGACACAGTTCTCTT 1575
2729 GACGCGCGGTCTCTCTGAAACACGACGCGTGGCGGCCCTTCTTTTGGGACACAGTTCTCTT 2670
1576 CAGCTGTCCACCTCTGCTGCTCTGAGTATATACAGGCGCTCGCGGACACCTCTAGTCCC 1635
2669 CAGCTGTCCACCTCTGCTGCTCTGAGTATATACAGGCGCTCGCGGACACCTCTAGTCCC 2610
1636 CTGATGCTCAGCACACGCTGACAGGATCACCTCTGGAACCGGACCTTGGGCGCACG 1695
2609 CTGATGCTCAGCACACGCTGACAGGATCACCTCTGGAACCGGACCTTGGGCGCACG 2550
1696 CCGGTGTGGAACCTCTACGCGGTCTCACCTTTCTACTGGCGCTCGAGGACCGCGGTGG 1755
2549 CCGGTGTGGAACCTCTACGCGGTCTCACCTTTCTACTGGCGCTCGAGGACCGCGGTGG 2490
1756 GCACACGGGGTGTCTCTCTCTAAACAGCAATGCGATGGTCTCTGAGCGGACGCTT 1815
2489 GCACACGGGGTGTCTCTCTCTAAACAGCAATGCGATGGTCTCTGAGCGGACGCTT 2430
1816 GCCCTTAGCTGAGGTGCACAGTGGGATCTCTGATGTCTACATCTTCTGCGGCGCACG 1875
2429 GCCCTTAGCTGAGGTGCACAGTGGGATCTCTGATGTCTACATCTTCTGCGGCGCACG 2370
1876 CCAAGAGCGTGTGCAGCAGTACCTGACGTTGTGGATACCCGTTCTATGCGGCGCATAC 1935
2369 CCAAGAGCGTGTGCAGCAGTACCTGACGTTGTGGATACCCGTTCTATGCGGCGCATAC 2310
1936 TGGGCGCTGGGCTTCCACCTGTGCGCTGGGCTACTCTCCACCGCTATCACCGGCCAG 1995
2309 TGGGCGCTGGGCTTCCACCTGTGCGCTGGGCTACTCTCCACCGCTATCACCGGCCAG 2250
1996 GTGGTGGAGACATGACAGGCGCCACTTCCCGCTGGAGTCC - AATGGAAAGCATGA 2054
2249 GTGGTGGAGACATGACAGGCGCCACTTCCCGCTGGAGTCCGAATGGAAAGCATGA 2190
2055 CTACATGGAATCCCGGAGGAGCTTCAACGTTCAACAGGATGGCTTCCGCGGACTTCCCGGC 2114
2189 CTACATGGAATCCCGGAGGAGCTTCAACGTTCAACAGGATGGCTTCCGCGGACTTCCCGGC 2130
2115 CATGTGAGGAGTGCACAGGCGCGCGCTTACATGATGATGATGATGATGATGATGATGATG 2174
2129 CATGTGAGGAGTGCACAGGCGCGCGCTTACATGATGATGATGATGATGATGATGATGATG 2070
2175 CAGCAGCTCGGCGCTGCGGAGCTACAGGCGCTACAGGAGGCTCTGCGGAGGGGGT 2234
2069 CAGCAGCTCGGCGCTGCGGAGCTACAGGCGCTACAGGAGGCTCTGCGGAGGGGGT 2010
2235 TTTTCATCAACAGAGACCGCGCTGATTTGGGAAGGATATGGCGCGGCTTCCACTGC 2294
2009 TTTTCATCAACAGAGATCGCGCAGCGCTGATTTGGGAAGGATATGGCGCGGCTTCCACTGC 1950
2295 TTTCCCGACTTCAACAAACCGCTTGGGAGGACATGGTGGGTGAGTT 2354
1949 TTTCCCGACTTCAACAAACCGCTTGGGAGGACATGGTGGGTGAGTT 1890
2355 CCATGACAGGTGCGCTTGGAGCGCTTGGATTTGACATGACAGGAGCTTCCAACTTCAT 2414
1899 CCATGACAGGTGCGCTTGGAGCGCTTGGATTTGACATGACAGGAGCTTCCAACTTCAT 1830
2415 CAGAGGCTCTGAGGAGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2474
1829 CAGGGGCTCTGAAACGGCT - CCCCACAACTAGAGCTGGAGAAACCAACTAGTACCTG 1771
2475 GGTGTGTGGGGGACCCCTCCAGGCGGCGCACCATCTGTGCTCTCAGCCACAGTCTCTCTC 2534
1770 AGTGTGTGGGGGACCCCTCCAGGCGGCGCACCATCTGTGCTCTCAGCCACAGTCTCTCTC 1711
2535 CACACACTACACCTGACAACTCTACGGCTGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2594
```

1710	Db	CACACACTAACCTGCACAACTCTTACGGGCTGACCGAGCCATCGCTCCCAACAGGGC	1651
2595	Qy	GCTGGTGAAGGCTCGGGGGACACGCCCAATTTGTGATCTCCGGCTCGACCTTTGCTGGCCA	2654
1650	Db	GCTGGTGAAGGCTCGGGGGACACGCCCAATTTGTGATCTCCGGCTCGACCTTTGCTGGCCA	1591
2555	Qy	CGGCCGATACGCCCGGCACCTGGACGGGGGACGTGTGAGCTCTCTGGGAGCAGCTCGCCTC	2714
1590	Db	CGGCCGATACGCCCGGCACCTGGACGGGGGACGTGTGAGCTCTCTGGGAGCAGCTCGCCTC	1531
2715	Qy	CTCCGTCGCAGAAATCTCTGCAGTTTAACTGCTCGGGGTGCCTCTGCTCGGGGCCGACGT	2774
1530	Db	CTCCGTCGCAGAAATCTCTGCAGTTTAACTGCTCGGGGTGCCTCTGCTCGGGGCCGACGT	1471
2775	Qy	CTCGGGCTTCTTGGGCAACACTCAGAGGAGCTGTGTGCGCTGGACCCACAGCTGGGGGC	2834
1470	Db	CTCGGGCTTCTTGGGCAACACTCAGAGGAGCTGTGTGCGCTGGACCCACAGCTGGGGGC	1411
2835	Qy	CTTCTACCCCTTATGCGGAACCAACACAGGCTCTGCTGCTGCCCCCAGAGCCGTAACG	2894
1410	Db	CTTCTACCCCTTATGCGGAACCAACACAGGCTCTGCTGCTGCCCCCAGAGCCGTAACG	1351
2895	Qy	CTTCAGGAGCGCGCCAGCAGGCGCATGAGAAAGCGCCTCACCTGCGCTACGCACTCCT	2954
1350	Db	CTTCAGGAGCGCGCCAGCAGGCGCATGAGAAAGCGCCTCACCTGCGCTACGCACTCCT	1291
2955	Qy	CCCCCACTCTACACACTGTTTCCACAGGCCACAGTCGCGGGGAGACCGTGGCCCGGCC	3014
1290	Db	CCCCCACTCTACACACTGTTTCCACAGGCCACAGTCGCGGGGAGACCGTGGCCCGGCC	1231
3015	Qy	CCTCTTCTGTGAGTTCCCAAGGACTCTAGGACCTGTGAGCAACACAGCTCTCTGTG	3074
1230	Db	CCTCTTCTGTGAGTTCCCAAGGACTCTAGGACCTGTGAGCAACACAGCTCTCTGTG	1171
3075	Qy	GGGGGAGCCTGCTCATCAACCCAGTGCTCCAGGCCGGGAAGCCCAAGTGAATGAGCTA	3134
1170	Db	GGGGGAGCCTGCTCATCAACCCAGTGCTCCAGGCCGGGAAGCCCAAGTGAATGAGCTA	1111
3135	Qy	CTTCCCCCTTGGGCACATGCTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCT	3194
1110	Db	CTTCCCCCTTGGGCACATGCTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCT	1051
3195	Qy	CCCAACCCCACTGCGAGCTCCCGTGAGCCAGCCATCCACAGCAGGGGAGTGGGTGAC	3254
1050	Db	CCCAACCCCACTGCGAGCTCCCGTGAGCCAGCCATCCACAGCAGGGGAGTGGGTGAC	991
3255	Qy	GCTGCGGGCCCCCTGGACACCATCAAGTTCACCTCGGGCTGGGTACATCATCCCCCT	3314
990	Db	GCTGCGGGCCCCCTGGACACCATCAAGTTCACCTCGGGCTGGGTACATCATCCCCCT	931
3315	Qy	GCAGGGCCCTGGCCTCAACCAACAGAGTCCCGCAGCAGGCCCATGGCTGTGGC	3374
930	Db	GCAGGGCCCTGGCCTCAACCAACAGAGTCCCGCAGCAGGCCCATGGCTGTGGC	871
3375	Qy	CCTAACCAAGGTTGAGAGGCCCGAGGGGAGCTGTTCTGGGAACATGGAGAGCCTGGA	3434
870	Db	CCTGACCAAGGTTGAGAGGCCCGAGGGGAGCTGTTCTGGGAACATGGAGAGCCTGGA	811
3435	Qy	AGTGCTGGAGGGAGGGGCTTACACACAGGTTCATCTTCTGGCCAGGAATTAACAGATCGT	3494
810	Db	AGTGCTGGAGGGAGGGGCTTACACACAGGTTCATCTTCTGGCCAGGAATTAACAGATCGT	751
3495	Qy	GAATGAGCTGTAGCTGTGACAGTGAAGGAGCTGGGCTGCGAGCTGCAG-AAAGTGACTG	3553
750	Db	GAATGAGCTGTAGCTGTGACAGTGAAGGAGCTGGGCTGCGAGCTGCAGAAAGTGACTG	691
3554	Qy	TCCTGGGCGTGGCCACGGCGCCACAGAGGTCCTCTTCCAAAGGTGTCCCTGTCTCCAAC	3613
690	Db	TCCTGGGCGTGG - CACGGCGCCACAGAGGTCTCTCTCAAAGGTGTCCCTGTCTCCAAC	632
3614	Qy	TCACCTACAGCCCCGACACCAAGGTCTTGACATCTGTGTCTCGCTGTTGATGGGAGAGC	3673
631	Db	TCACCTACAGCCCCGACACCAAGGTCTTGACATCTGTGTCTCGCTGTTGATGGGAGAGC	572

3674 AGTTTCTGTCAGCTGTT 3694
 ||||||||||||||||
 571 AGTTTCTGTCAGCTGTT 551

RESULT 8
 AF171665 2814 bp mRNA linear MAM 04-JUL-2000
 LOCUS Bos taurus acidic alpha-glucosidase mRNA, complete cds.
 DEFINITION
 AF171665
 VERSION AF171665.1 GI:8925837
 KEYWORDS
 SOURCE Bos taurus (cow)
 ORGANISM
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 1 (bases 1 to 2814)
 Dennis,J.A., Moran,C. and Healy,P.J.
 The bovine alpha-glucosidase gene: coding region, genomic
 structure, and mutations that cause bovine generalized glycosgenosis
 Mamm. Genome 11 (3), 206-212 (2000)
 JOURNAL
 MEDLINE 20188772
 PUBMED 10723725
 REFERENCE
 2 (bases 1 to 2814)
 Dennis,J.A., Moran,C. and Healy,P.J.
 Direct Submission
 Submitted (22-JUL-1999) Agriculture, Elizabeth Macarthur
 Agricultural Institute, Woodbridge Road, Menangle, New South Wales
 2568, Australia
 FEATURES
 Location/Qualifiers
 1..2814
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /cell_type="fibroblast"
 1..2814
 /note="AAG"
 /codon_start=1
 /product="acidic alpha-glucosidase"
 /protein_id="AAF81636.1"
 /db_xref="GI:8925838"
 /translation="MMRWSPCPPLGVCTLLSLLGLHLLHLEVVPELRFSGFOD
 EIHQACQGNASSPECEGSRRAAPTQCDLPNSRPDCAPDKGITPQCEARCGCYMPAE
 WPDARQGWQCFPPPSYSEYRLENLTITGTATLRAVTPFFPKDINLLDMLM
 ETESLHFTKCPDIPANRYEVLPTPRVYEQAPFTLYSEFSEBPPGVVRRKLDGRVL
 LNTSPHFDADQLQSLTSQHITGLAHLGSLMSTNWTKITLNNRDIAPEPV
 LNYGSHFVLYLDGGLHGFVLNNSAMDVLQSPALSWRSTGGTLDVYIFLGPPEP
 KSVYQQLDVGVPPVPPVWGLGFLHRCWGYSTAITROVVNMTRAYFLPDVQWNDL
 DYMDARDPTFNKHDGDPAPMVQELHQGBRYVIMVDPAISSSGPGATVRPYDEGLR
 RGVFTINERGQPLIGQWFGLTAPDFTNPETILDWQDNVTEPHAQVFPDGMIDMNE
 SNFNVSGSCDPSLENNPPYLPVGVGGTLRAATICASHQFLSTHYDLDHNLGLUTE
 ALASHRALVKARGMFPVTSRSFTAGHRYSGHWTGDVMSNWEQLSYSVPEILLFLNL
 GVLPGVADICGFTNLSBELCVRTQLGAPYPMRNHNLNSQPPYRFSYAQQAM
 RKAFATVYLLPLYTLFRAHVRGETVARPLFEPDPTSTVQRQLLWGALLIT
 PVLEAKVEVTFVPGTVDLOTVPMEAFGSLPPPAPLTSVTHSGOWVTLSAPLDT
 INVHLRAGHILPMQGPALITTESRKHOMALAVALTASGEAQGELFWDDGSLVLDGG
 DYTQLIFLAKNNFTFNKLVRVSESGASLQURNVTLGVATAPQVLCNSVPSVSNFTFS
 PDTETLAIPLVSLTMGEGFVISM"

ORIGIN
 Query Match 54.9%; Score 2033.2; DB 4; Length 2814;
 Best Local Similarity 86.0%; Pred. No. 2.6e-305;
 Matches 2280; Conservative 0; Mismatches 363; Indels 9; Gaps 2;

QY 1045 GCACACCCCGCGTCCACAGCAGTGTCCACACAGTGCAGCTGCCCCCCCAACAGCGGC 1104
 Db 172 GAATGCCGTGGCAGCCCGCAGCGCAGCACCTACACAGTGCAGCTGCCCCCCCAACAGCGGC 231

QY 1105 TTGATTCGCCCTGACAGGCCATCACCACGACAGTGCAGGCCCGCGGCTGCTGC 1164
 Db 232 TTGACTGCGCCCCAGCAGGCGCATCACCCCGCAGCGTGCAGGCCCGGCTGCTGC 291


```
Db      2443  GCCTTACGACACAGAGTCCGCAAGACGACACATGCGCTGTGGCCCTGACAGCC 2502
Qy      3385  GTGGAGAGCCGAGGAGCTGTCTGGAGAGTGGAGAGCCTGGAAGTCTGGAG 3444
Db      2503  AGTGGAGAGCCGAGGAGCTGTCTGGAGAGTGGAGAGCCTGGAGTCTGGAT 2562
Qy      3445  CGAGGGCCCTACACACAGGTCATCTTCTGGCCAGGAATAACACGATCGTGAATGAGCTG 3504
Db      2563  GGTGGGAGCTACACACAGCTCATCTTCTGGCCAGGAACAACACCTTTGTAAACAAGCTG 2622
Qy      3505  GTACGTGTACAGTACGAGGAGTGGCTGACCTGACCTGACGACGAGTGTCTCTGGGCGTG 3564
Db      2623  GTGACAGTGTGAGTGTGAGGAGGAGCCTGCTGACCTGAGGAATGTGACCGTCTGGGCGTG 2682
Qy      3565  GCCACGCGCCGACGAGGTCCTTCCAAAGGTCCTGCTGCTGCTCCAACTTCACTACAGC 3624
Db      2683  GCCACAGCCCCACGAGGTCCTTCCAAAGGTCCTGCTGCTGCTCCAACTTCACTACAGC 2742
Qy      3625  CCGACACACAGGTCCTGAGCATCTGTGCTGCTGCTGCTGATGGAGAGCATTTCTCGTC 3684
Db      2743  CCGTACACAGAGACTCTGGCCATCCCTGCTGCTGCTGCTGACCATGGAGAGCATTTGTGATC 2802
Qy      3685  AGCTGTGTGA 3696
Db      2803  AGCTGTGTGA 2814

RESULT 9
BC061753
LOCUS
DEFINITION
Rattus norvegicus glucosidase, alpha; acid (Pompe disease, glycogen storage disease type II), mRNA (cdna clone MGC:72625 IMAGE:5599964), complete cds.
ACCESSION BC061753
VERSION GI:38197415
KEYWORDS
SOURCE MGC.
ORGANISM Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3408)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Utskin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.W., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.M., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettaman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krywinski,M.I., Skalski,U., Smallos,D.B., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
TITLE 2 (bases 1 to 3408)
AUTHORS Strausberg,R.
JOURNAL Direct Submission
Submitted (05-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
```

COMMENT

Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Jeff Green/Paturu Kondatiah, NCI.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter,N., Avale,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghghi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Teurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 137 Row: i Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES
source

1..3408
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="MGC:72625 IMAGE:5599964"
/tissue_type="Prostate, pool of NCI_CGAP_Pr30, 40, 41 and NCI_CGAP_Pr29, 39, 42"
/clone_lib="NCI_CGAP_Pr50"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6.1"
1..3408
/gene="Gaa"
/note="synonym: MGC72625"
/db_xref="LocusID:367562"
/db_xref="RGD:735227"
171..3032
/gene="Gaa"
/codon_start=1
/product="glucosidase, alpha; acid (Pompe disease, glycogen storage disease type II)"
/protein_id="AAH61753.1"
/db_xref="GI:38197416"
/db_xref="LocusID:367562"
/db_xref="RGD:735227"
/translation="WNIRKPLCSNVGACTLVSLTAVILGHLMLRELMLLPQDLHE SSSGLWKTYPHQSYPAPLHIOEHAOLRAVPTQCDVTPNSRFDCAKPKITQEQ CEARCCWVPAQVNGVPMGQPCWCFPPSPYSYRLNLSSTESGYTALTRISPTFF PKDVLTLQLEVLMTDSRLHFMKDPTSKRYEVLPLETLPVLQAPLQATTCASSHQFLS GIVVRKGLGRVLTNTVAPFPADQFLQSLSPSHQTAGLGHLSPLMLSTWTRIT TLMNRDVAPOGCVNLYGSHFPFLALEDGLAGHFLNSNAMDVLPQSPALTWRTGT GLDVYVFLPSPKSVQVLDVGVPPMPYVGLGFLCRWGYSTAVRQVENVNT RTHFLDQVMDLDYMDARRDFTNQDGFADPMVHELHQGRRYMMIVDPAISSSG PASTRTRDEGLRRGVFTINETGQLIKGWEGSTAFDFTINPDLQWQMVSEPHA QVPPGMWIDMNPENFIRGSOQCGPDNELENPPYVGVGALQATTCASSHQFLS THYNILNYGLTEALIASRALVTKTRTPFVSRSTFAGRYAGHTGDTVMSSWHL AYSVPEILQNLGLPVGADICGFGQNTTELCLVRLTGALGFYPMRNRNDLNSLPQ EYRSETAQQAAMKAFPLRVALLPVLTLFHGAHVKGDTVARPLFLEPPDPSTWSV DRQLLNGPALLITPVLEKGTDTGCFPKGMWYNLQMPVETLGSIPSSSPASSGSI VHSKQWLTEAPLDTINHLRAGYIIPLQGPSTLTTERKQKQPMALAVLTSGEASG ELFWDDGESLGLVREGATVLTFSKNTINVKLVHVTKGEGELQRLREVTILGVTAP TQVLSNGISVSNFTYSPDDKSLISIPVSLMGERFQIDWS"

gene

CDS

ORIGIN

Query Match 51.4%; Score 1903; DB 10; Length 3408;
Best Local Similarity 82.7%; Pred. No. 3.7e-285;
Matches 2189; Conservative 0; Mismatches 455; Indels 3; Gaps 1;

QY	1048	CACCCCGCGCGTCCAGACAGTGGCCACACAGTGCAGCTCCCTCCCAACAGCGCTTC	1107		Db	1461	CTCCACCAAGGGTGGCGCGGTACATGATGATTGTGGACCTCGCCATCAGCAGTTCAGGC	1520	
Db	381	CACGCCAGACAGCTAAGACAGTATACCAACAGTGTGATGTAACCCCAACAGCGCTTT	440		QY	2198	CCTGCCGGAGCTACAGGCCCTACACGAGGGTCTCGCGAGGGGGTTTTTCATCACCAC	2247	
QY	1108	GATTTGCCCTTGACAAGGCATACACCAAGGAACAGTGGAGGCCCGCGCTGTGCTAC	1167		Db	1521	CCTGTGGAGTTACAGGCCCTATGATGAGGGTCTTCGAGGGGGGTGTTTCATCACCAT	1580	
Db	441	GACTGTGCCCGGACAAAGGCATACCCAGGAGCAGTGGCGGCCGAGGGCTGCTGCTGG	500		QY	2248	GAGACCGGCCAGCCCTGATTGGGAAGGTATGGCCCGGGTCCACCTGCCCTTCCCGCATTC	2307	
QY	1168	ATCCCTGCAAAACAGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGGTGTCTTCCCA	1227		Db	1581	GAGACTGGGCAGCATTGATTGGGAAGGTTTGGCCCTGGATCCACGCCCTTCCCTGATTTC	1640	
Db	501	GTCCCAAGCAGGAGGTGCTGTAATGGGCCAGTGATGGGGCAGCCCTGGTGTCTTCCCT	560		QY	2308	ACCAACCCCAACAGCCCTGGCTGGTGGAGAGACATGGTGGCTGAGTTCATGACACAGGTG	2367	
QY	1228	CCAGCTATCCCAAGCTACAAGCTGGAGAACCTGAGCTCCTCTGAATGGGCTACAGGCC	1287		Db	1641	ACCAACCCCTGAGACCCCTCGACTGGTGGCAGGACATGGTGTCTGAGTTCATCCTCCAGGTG	1700	
Db	561	CCAGCTATCCCAAGCTACCGCTAGAGAACCTGAGCTCTACAGAGTCAGGGTACACAGCC	620		QY	2368	CCCTTCAGCGGCTGTGGATTGACATGAACGAGCCCTCCAACTTCATCAGAGGCTCTGAG	2427	
QY	1288	ACCTGACCGGTACCAACCCGACCTTCTTCCCAAGGACATCCTGACCTCGGGCTGGAC	1347		Db	1701	CCCTTCGATGGCATGTGGATCGACATGAACGACCGTCCAACTTCATCAGAGGCTCTCAG	1760	
Db	621	ACCTGACCGGACCAAGTCCGACCTTCTTCCAAAGGATGTGTGACCTTACAGCTGGAG	680		QY	2428	GACGGCTGCCCAACAATGAGCTGGAGAACCCACCCCTAGTGCCTGGGGTGGTGGGGG	2487	
QY	1348	GTGATGATGAGACTGAGAACCGCCCTCCACTTCACGATCAAGATCCAGCTAACAGCGC	1407		Db	1761	CAGGGCTGCCCTGACAATGAACCTAGAGAACCCCTTATGTGCCCGGGTGGTGGCGGG	1820	
Db	681	GTGCTATGAGACTGACAGCGCCCTCCACTTTATGATCAAGATCCCACTAGTAGCGC	740		QY	2488	ACCTTCAGGGCGGCCACCATCTGTGCTCCAGCCACAGTTTCTCTCCACACACTACAC	2547	
QY	1408	TACGAGGTGCCCTTGGAGACCCCGCGTGTCCAGACCGGGCACCGTCCCACTCTACAGC	1467		Db	1821	GCTTTCAGGGCGGCCACCATCTGTGCTCCAGCCACAGTTCCTCTCTACACACTACAC	1880	
Db	741	TACGAAGTGCCCTGGAGACCCCGCGTGTCTTGAGCCAGCGGCATCCCCCACTTACAGT	800		QY	2548	CTGCACAACTCTACGGCTGACCGAAGCATCGCTCCACAGGGCGCTGGTGAAGGCT	2607	
QY	1468	GTGGAGTTCTCCGAGAGCGCTTCGGGGTGATGTGTCAACGGCAGCTGACCGCCCGGTG	1527		Db	1881	CTCCAAACCTTATCGGCTCACTGAGCAATGCTCCAGCAGGGCCCTAGTCAAGACT	1940	
Db	801	GTGGAGTTCTCAGAGAGCGCTTCGCGGTGTGCTTCGAGGAAGCTTGGTGGCGAGTG	860		QY	2608	CGGGGACACGCCCAATTTGTGATCTCCGCTCGACTTTGTGGCCACGGCCGATACGCC	2667	
QY	1528	CTGCTCAACAGAGGTGGCGCCCTGTCTTTGGGACAGTTCCTTCAGCTGTCCACC	1587		Db	1941	CGGGGAACAGGACCTTTGTGATCTCCGCTCAAACCTTTGGGGCCACGGCCGGTATGCC	2000	
Db	861	CTGCTCAACACACCGTGGCGCCCTCTTCTTCGCTGACAGTTCCTGACAGTGTCCACC	920		QY	2668	GGCCTAGGACGGGGGACGTGTGGAGCTCTGGGAGCAGCTCGCTCTCCCTCGCCAGAA	2727	
QY	1588	TGCTGCCCTCGCAGTATATCAGAGCCTGCGGAGCACTCAGTCCCCTGATGTCAGC	1647		Db	2001	GGTCACTGGACAGGGGACGTGTGGAGCTCTTGGAGCATCTTGGGAGCATCTGTGCCAGAA	2060	
Db	921	TCACTGCCCTCCAGCACATGGCAGGCTGGGTGAGCACCTCAGTCCACTCATGCTCAGC	980		QY	2728	ATCTCTGCAAGTTTAAACCTCTGGGGGTGCTCTGTGTCGGGGCCGAGCTCTGCGGCTTCCTG	2787	
QY	1648	ACCAGCTGGACAGGATACCTGTGGAAACCGGACCTTGGCCCAACGCCCGGTGGAC	1707		Db	2061	ATCTCTGCAAGTTTAAACCTGTGTCGGCGTGCCTTGGTGGGGCAGACATCTGCGGCTTCAG	2120	
Db	981	ACCGAATGGAAGTGCATCACTCTCTGGAACCGGGATGTGGCACCTCGCAAGGTGTCAAC	1040		QY	2788	GGCAACACTCAGAGAGCTGTGTGCTGGACCCAGCTGGGGGCTTCTACCCCTTC	2847	
QY	1708	CTTACGGGTCTCACCCTTTTACCTGGCGTGGAGAGCGGGGTGGGACACAGGGGTG	1767		Db	2121	GGAAACACACAGAGAAGAGCTGTGTGCGCTGGACCCAGTTGGGGGCTTCTACCCCTTC	2180	
Db	1041	CTCTATGGGTACACCTTTTACCTTGGCACTGGAGGACGGTGGCTTGGCTCAOGGTGTC	1100		QY	2848	ATCGGGAACCAACACAGCTGCTCAGTCTGCCCCAGGAGCCGTACAGCTTCAGCGAGCG	2907	
QY	1768	TTCCTGCTAAACAGCAATGCCATPGATGTGCTCGACCGGAGCCCTGCTTAGCTGG	1827		Db	2181	ATGCGGAACCAACAATGACCTGAATAGTTCCTCAGGAGCCGTACAGGTTTACGCGAGAG	2240	
Db	1101	TTCCTGCTAAATAGCAATGCCATGGATGTGCTCTGACGCCACAGCCCGCTTAACTTGG	1160		QY	2908	GCCCAAGAGCCATGAGGAAGGCCCTCACCTTCGCTACGCACTCTCCTCCCCACCTCTAC	2967	
QY	1828	AGTTCAGAGTGGGATCTCGGATGTATCATCTCTGGGCCAGAGCCCAAGAGCGTG	1887		Db	2241	GCACAGAGCCCATGAGGAAGCCCTTACCTTTCGCTATGCTTCTACCTTACCTGTAT	2300	
Db	1161	AGGTCAACAGGCGGAATCTCGAATGTTTATGTGTTCTTAGGCCACAGCCCAAGAGTGT	1220		QY	2968	ACACTGTTCCACAGGCCCAACGTGCGGGGGAGACCGTGGCCCGGCCCTCTTCTCTGGAG	3027	
QY	1888	GTGACAGTACCTTGCAGCTTGTGGGATACCCGCTTATGCTCCGCACTACTGGGGCTGGG	1947		Db	2301	ACACTCTTCCAGCGGCCCAACGCTCAAAGGGGACAGGTTGGCCCGGCCCTCTTCTCTGGAG	2360	
Db	1221	GTCCAAACATACCTGATGTGCTGGGATACCCCTTATGCTCCGCTAGTGGGGCTGGGC	1280		QY	3028	TTCCCAAGGACTCTAGCACCTGAGTGTGGACCAACAGCTCTCTGTGGGGGAGGCCCTG	3087	
QY	1948	TTCCACCTGTGCGCTGGGGCTACTCTCTCAACGCTATACCCCGCAGGTGTGGAGAAC	2007		Db	2361	TTCCCTTGAGGATTCACAGCACCTGGTGTGTGGACCGCAGCTCTTGTGGGGGGCGGCCCTG	2420	
Db	1281	TTCCACCTGTGCGCTGGGGCTACTCTCTCACCGCCATCGTCCGACAGGTTGTGGAGAAC	1340		QY	3088	CTCATCACCCCAAGTGTCTCCAGGCCCGGAAGGCCAGTGAAGTGAAGTCTTCCCTTTGGGC	3147	
QY	2008	ATGACAGGGCCCACTTCCCTCGAGCTCCATGGAACGACCTGACTACATGGAATCC	2067		Db	2421	CTTATCACACCTGTGCTTGAACCTGGGAAAACCTGACGTGACAGGCTACTTCCCAAGGGC	2480	
Db	1341	ATGACAGGACGACACTTCCCTCGAGCTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGT	1400		QY	3148	ACATGATGACGACTGTGACAGCGGTGCCAATAGAGGCCCTTGGAGCCCTCCACCCCACT	3207	
QY	2068	CGAGGAGACTTCACTGTTCAACAGGATGGCTTCCGGGACTTCCCGGCCATGTGTGACGAG	2127		Db	2481	ATGTGATCAACCTTGAGATGGTACCTGTGGAGACCCCTTGGTAGCCTCTTCTTCTGCA	2540	
Db	1401	CGAGAGACTTCACTTCAACAGGACGGCTTTCAGAGACTTCCAGACATGTGTGACGAG	1460		QY	3208	GCAGTCTCCCGGTGAGCGACCA---TCCACAGCAGGGGCGAGTGGGTGACGCTCGCGGC	3264	
QY	2128	CTGCAACAGGGCGCGCGCTACATGATGATCGTGGATCCTGCCATCAGCAGCTCGGC	2187						

Db 435 CAGACTGAACAGCCCAAGAACACCCACACAGTGTGATGTGCCCCCAGCAGCGCTTT 494
QY 1108 GATTGGCCCTTGACAAGGCATCACCCAGGAACAGTGCAGGCCCGCGCTGCTGTACT 1167
Db 495 GACTGTGCCCCGACAAAGGCATCTCACAGGAGCAATGAGAGCCCGCGCTGTGCTAT 554
QY 1168 ATCCCTGCAAAAGCAGGGGTGACGAGAGCCACAGATGGGCGACGCCCTGTCTTCCCA 1227
Db 555 GTCCACAGCAGGCGAGGTGTGAAGGAGCGCAGATAGGCGAGCCCTGTGTCTTCCCT 614
QY 1228 CCCAGTACCCAGCTACAGCTGAGGAGACCTGAGCTCTCTGAAATGGGCTACACGCC 1287
Db 615 CCCAGTACCCAGCTACCGTCTAGAGAACCTGAGCTCTACAGAGTCGGGGTACACAGCC 674
QY 1288 ACCCTGACCGTACCAACCCACCTTCTTCCCAAGGACATCTGACCCCTGGGCTGGAC 1347
Db 675 ACCCTGACCGGTACAGCCCGACCTTCTTCCCAAGGATGTGTGACCTTACAGCTGGAG 734
QY 1348 GTGATGATGGAGCTGAGAACCGCTTCCACTTCAAGATCAAAAGATCCAGCTAACAGGCGC 1407
Db 735 GTGCTGATGGAGACAGACAGCGCGCTTCCACTTCAAGATCAAAAGATCCTGTAGTAAGCGC 794
QY 1408 TAGGAGTGCCTTGGAGACCCCGGTGTCCACAGCCGGGACCCGTCGCCACTTACAGC 1467
Db 795 TAGGAAGTGCCTTGGAGACCCCGGTGTGCTGAGCCAGGCACTATCCCACTTTACAGC 854
QY 1468 GTGGAGTCTCCGAGGAGCCCTTGGGGGTGATGTCACCGGAGCTGACCGCGCGCTG 1527
Db 855 GTGGAAATCTCAGAGGAAACCTTTGAGTGTGCTGTAGGAGCTTGGTGGCCGAGTG 914
QY 1528 CTGCTGAACACAGCGGTGGCGCCCTGTCTTTGGGACCAAGTTCCTCAGCTGTCCACC 1587
Db 915 TTGCTGAACACAAACCGTGGCGCCCTGTCTTCCGCTGACAGTTCCTGAGCTGCCACT 974
QY 1588 TCGCTGCCCTCGCAGTATATACAGGCGCTGCGGAGCACTCAGTCCCCTGTAGCTCAGC 1647
Db 975 TCCCTGCCCTCCAGACATCATCAGGCGCTGGGGGAAACACTCAGGCCACTCATGTCTCAGC 1034
QY 1648 ACCAGCTGCACAGGATCAACCTGTGGAACCGGGAGCTTGGCGCCACGCGCCGTGCGAAC 1707
Db 1035 ACCGACTGGGCTGTATCACCTCTTGGAAACCGGGACACACCACTCGCAAGTACCAC 1094
QY 1708 CTCTACGGGTCTCACCTTTTCTACTTGGCGCTGGAGGACGGGGGTGCGCAACCGGGTG 1767
Db 1095 CTCTACGGGTCTCATCTTTCTACTTGGCACTGGAGGACGGTGGCTTCACTGCTGCTGTC 1154
QY 1768 TTCTCTGCTAAACAGCAATGCCATGGATGGTCTGTCAGCGCGAGCCCTGCTTACTGCTG 1827
Db 1155 TTCTTCTGCTAAACAGCAATGCCATGGATGTCATCTGCAACCCAGCGCCCTTAACCTGG 1214
QY 1828 AGGTGACAGGTGGGATCTCTGATGTCTATCTTCTGGGCGCAGAGCCCAAGAGCGTG 1887
Db 1215 AGGTCAACGGGCGGGATCTTGGATGTATGTCTTAGGCCCGCAGAGCCCAAGAGCGTT 1274
QY 1888 GTGACAGTACTGTGACCTTGTGGGATACCCGTTTCACTGCGGCATPACTGGGGCTTGGG 1947
Db 1275 GTGCAACAATACCTGTGATGTGGGATACCCCTTCACTGCTCCATPACTGGGGCTTGGG 1334
QY 1948 TTCCACTGTGCGCTGGGGCTACTCTCCAGGCTATCACCGCATCACCGCAGGTGGTGGAGAC 2007
Db 1335 TTCCACTGTGCGCTGGGGCTACTCTCCAGCGCATTTGTCCGCGAGGTAGTGGAGAAC 1394
QY 2008 ATGACAGGGGCCACTTCCCTTGGAGCTTCCAAATGGAACGACTGGACTACATGGACTCC 2067
Db 1395 ATGACAGGACACACTTCCGCTGGATGTGCAATGGAATGACTGACTTACATGAGCGCC 1454
QY 2068 CGAGGGACTTCACTGTTCAACAGGATGGTTCGGGATTTCCGGGCAATTCGGGCGCATGTGCAAGGAG 2127
Db 1455 CGAAGAGACTTCACTTCAACAGGACAGCTTTGCCGACTTCCAGACATGTGTGGGGAG 1514
QY 2128 CTGACAGGGCGCGGCTTACATGATGATGCTGGATCTTCCGCTATCAGCAGCTCGGCG 2187
Db 1515 CTGACAGGGGTGGCGGCTTACATGATGATGCTGGACCTTCCATCAGCAGCGCAGGC 1574

QY 2188 CCTGCGGGAGCTACAGCCCTTACGACGAGGCTCTGCGAGGGGTTTTCATCACCAAC 2247
Db 1575 CCTGCTGGAGTTACAGGCCCTACGACGAGGCTCTGCGAGGGGTGTTCATCACCAAC 1634
QY 2248 GAGACGGGACAGCCCTGATTTGGGAAAGGTATGGCCCGGGTCCACTGCCCTTCCCGACTTC 2307
Db 1635 GAGACTGGGACAGCCCTGATTTGGGAAAGGTATGGCCCGGGTCCACTGCCCTTCCCTGATTTTC 1694
QY 2308 ACCAACCCACAGCCCTGCGCTGTGGGAGGACATGTGGCTGAGTTCTCATGACAGGCTG 2367
Db 1695 ACCAACCTTGAGACCCCTTGAATGGTGGCAGGACATGTGTCTGAGTTCCAGCCCGAGGTTG 1754
QY 2368 CCTTTCGACGGCTTGTGATTTGACATGAACGAGCCCTTCCAACTTTCATCAGAGGCTCTGAG 2427
Db 1755 CCTTTCGATGGCATGTGGCTCGACATGAACGAAACGTCCTCAACTTTCGTTAGAGGCTCTCAG 1814
QY 2428 GAGGCTGCCCAACAATGAGCTGAGAACCCACCTAGCTGCCCTGGGGTGGTGGGGG 2487
Db 1815 CAGGGCTGCCCAACAATGAACTGGAGAACCCCTTATGTGCCCGGGGTGTGTGGCGGG 1874
QY 2488 ACCCTCCAGGGGCGCACCATCTGTGCTCCAGCCACCACTTCTTCTCCACACACTACAAAC 2547
Db 1875 ATCTTGCAGGCGAGCCCATCTGTGCTTCCAGCCACCAATTCCTCTCCACACACTACAAAC 1934
QY 2548 CTGCAACAACCTTACGGCTTGACCGAAGCCATTCGCTCCACAGGGCGCTGTGAAGGCT 2607
Db 1935 CTCCACAACCTGTACGGCTTCACTGAAGCTATCGCTCCAGCAGGGCGCTGTGTCAAGACT 1994
QY 2608 CGGGGAGACGCGCCATTTGTGATCTCCCGCTCGACTTGTGCTGGGCAACGGCGGATACGCC 2667
Db 1995 CGGGGAAACGACCCCTTTGTGATCTCCCGCTCAACCTTCTCGGGGCAACGGCGGACTCGCT 2054
QY 2668 GGCACCTGACCGGGGACGCTGTGGAGCTCTGGGAGCAGCTCGCTCTCTCCGTGCAGAA 2727
Db 2055 GGTCACTGACAGGGGATGTGCGGAGCTCTTGGGAGCATCTTGCACTCTCTGTGCCAGAC 2114
QY 2728 ATCTCTGCAATTAACCTCTGCGGGTGTCTTGTGTGGGGCGACGCTGTGCGGCTTCTGTG 2787
Db 2115 ATCTCTGCAATTAACCTCTGCGGGTGTCTTGTGTGGGGCGACATCTGCGGCTTCATA 2174
QY 2788 GGCACACCTCAGAGGAGCTGTGTGCTGGAGCCAGCTGGGGGCTTCTTACCCCTTC 2847
Db 2175 GGAGACAGCTCAGAAAGAGCTGTGTGCTGGAGCCAGCTTGGGGGCTTCTTACCCCTTC 2234
QY 2848 ATGCGGAAACCAACAGACCTGCTCAGTCTGCGCCAGGAGCCGTACAGCTTTCAGCGAGCG 2907
Db 2235 ATGCGGAAACCAACATGACCTGATAGCTGCTCTGAGGAGCCGTACAGGTTTCAGCGAGAG 2294
QY 2908 GCGCAGAGGCGCATGAGAAAGCCCTCACTGCTGCGCTAAGCAGCTCTCTGCGGAGGAGCCCTG 3087
Db 2295 GCGCAGAGGCGCATGAGAAAGCCCTTCGCTTACGCTATGCGCTTCTGCGCTTACCTGTAC 2354
QY 2968 ACACGTTCCTCAGGAGCCAGCTGCGGGGAGACCGTGGCCCGGCGCTTCTTCTGGAG 3027
Db 2355 ACTCTCTTCCACCGCGCCCACTGTCAGAGGAGACAGGTTGGCGGCGCTTCTTCTGGAG 2414
QY 3028 TTCCCAAGAGGACTTACGACCTGTGAGCTGTGGACCAACAGCTCTGTGGGGGAGGCGCTG 3087
Db 2415 TTCCCTTGAGGATCCAGACACCTGTGTGTGACCCGCGAGCTCTTGTGGGGCGGCGCTG 2474
QY 3088 TTTCATCAACCCAGTCTCCAGCGCGGAGGCGGAGTGTGCTGTACTTTCCTTGGGCG 3147
Db 2475 CTTCATCAACCTGTCTTGAGCTTGGGAAACTGAAGTGAACGGGCTTCTTCCCAAGGCG 2534
QY 3148 ACATGGTAGGACCTCAGACCGTGCMAATAGAGGCGCTTGGCAGCTTCCACCCACCT 3207
Db 2535 ACGTGGTCAACATCAGAGTGTGTGAGTGGATTCCCTCGGTACTCTCCCTTCTCCATCA 2594
QY 3208 GCAGCT---CCCGGTGAGCCAGCCATCCACAGCAGGGGCGAGTGGGTGACGCTGCGCGGC 3264
Db 2595 TCGGCTTTCATCTTTCAGATCTGCTCTCCAGACAGGGGCGAGTGGCTGACACTGGAGGCC 2654


```
QY 3265 CCCTGGACACCATCAACGTCACCTCCGGGCTGGGTACATCATCCCTCGAGGCGCT 3324
Db 2655 CCACTGGATACCATCAACGTCACCTCGAGGAGGGGTACATCATCCCTCGAGGCTCC 2714
QY 3325 GGCTCACAACACACAGAGTCCCGCCAGAGCCCATGGCCCTGCTGGCCCTTAACCAAG 3384
Db 2715 AGCTCACAACACACAGAGTCCCGAAGACGCCCATGGCTCTGGCTGGCATTAACAGCA 2774
QY 3385 GTGGAGAGGCCCGAGGGAGCTGTTCTGGGACGATGGAGAGAGCTGGAAAGTCTGGAG 3444
Db 2775 AGCGGCGAGGCGGATGGGAGCTGTTCTGGGACGAGCGGAGAGCTTGCAGTCTCGGAG 2834
QY 3445 CGAGGGCCCTACACACAGCTCATCTTCTGGCCAGGAATAACACGATCGTGAATGACTG 3504
Db 2835 CGTGGGGCTTACACATGTGTCACTTCTCAGCCAAAGAAACAATACCAATGTGAACAAGTTA 2894
QY 3505 GTACGTGTACACAGTGAAGGAGCTGCGCTGACGCTCAGAGGTGACTGTCTGGGCGGTG 3564
Db 2895 GTGCGTGTACACAGAGGAGGAGCTGAGCTTACACTGAGGAGGTGACCGTCTTGGAGTG 2954
QY 3565 GCCACGGCCCGCAGCAGGTCTCTCCAAAGGTGTCCTGTCTCCAACTTCACTACAGC 3624
Db 2955 GCCACAGCTCTTACCAGGTCTCTTCCAAAGGATCCCTGTCTCCAACTTCACTACAGC 3014
QY 3625 CCGACACCAAGTCTCTGACATCTGTGCTGCTGCTGCTGATGGAGAGAGTTTCTCGTC 3684
Db 3015 CTTGACAAAGAGCTGGCCATCTCTGCTCACTGCTGATGGAGAGCTGTTTCAAATC 3074
QY 3685 AGCTGCT 3691
Db 3075 AGCTGCT 3081

RESULT 11
AX305670
LOCUS AX305670 3364 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 421 from Patent WO0188188.
ACCESSION AX305670
VERSION AX305670.1 GI:17645110
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1. Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 421 22-NOV-2001;
SCHOOL Juridical Person Nihon University (JP)
FEATURES
source
1. .3364
/organism="Mus musculus"
/mol_type="unassigned DNA"
/db_xref="taxon:10090"

ORIGIN
Query Match 50.8%; Score 1881.4; DB 6; Length 3364;
Best Local Similarity 82.1%; Pred. No. 8.2e-282;
Matches 2174; Conservative 2; Mismatches 468; Indels 3; Gaps 1;

QY 1048 CACCCCGGCGTCCAGAGCAGTGCCACACAGTGCAGCGTCCCCCCCCAACAGCGCTTC 1107
Db 323 CAGACTGAACAGGCCAAAGAACACACACAGTGTGTGCGCCCGCCAGCAGCGCTTT 382
QY 1108 GATTGCGCCCTGACAGGCCATCACCCAGGAACAGTGGAGGCCCGCGGCTGCTCTAC 1167
Db 383 GACTGTGCCCCACAAAGGCATCTCACAGGAGCAATGCGAGGCCCGAGGCTGCTCTAT 442
QY 1168 ATCCCTGCAAGCAGGGCTGCGAGGAGCCAGATGGCGAGCGCTTCTTCTTCCCA 1227
Db 443 GTCCAGAGGGGAGGTGCTGAAGAGGACACAGATAGGGCAGCCCTTCTTCTCCCT 502
QY 1228 CCAGCTACCCAGCTACAGCTGGAGAACCTGAGCTCCTCTGAAATGGGCTACACGGCC 1287
```

```
Db 503 CCCAGCTACCCCAAGCTTACCGTCTAGAGAACCTGAGCTCTACAGAGTCGGGGGTACACAGCC 562
QY 1288 ACCCTGACCCGTACCAACCCCACTTCTTCCCAAGGACATCTTGACCCCTCGGGCTGGAC 1347
Db 563 ACCCTGACCCGTACCAACCCCACTTCTTCCCAAGGATGCTGACCTTACAGCTGGAG 622
QY 1348 GTGATGATGGAGATGAGAAACCGCTCTCACTTCAAGTCAAAAGATCCAGCTAACAGGCGC 1407
Db 623 GTGCTGATGGAGACAGACAGCGCCCTCCACTTCAAGATCAAAAGATCCCAAGTAAAGCGC 682
QY 1408 TAGGAGGTGCCCTTGAGAGACCCCGGTGTCACAGCCGGGACCGCTCCCACTCTACAGC 1467
Db 683 TAGGAGGTGCCCTTGAGAGACCCCACTGCTGAGCCAGGACCACTCCCACTTACAGC 742
QY 1468 GTGGAGTTCTCCGAGGAGCCCTTCCGGGGTGTGTCGACACCGGAGCTGGAACGGCCCGTG 1527
Db 743 GTGGAATTCTCAGAGGAAACCCCTTGGAGTGTGCTTCTGAGGAAGCTTGGTGGCCGAGTG 802
QY 1528 CTGCTGAACACAGCGGTGGCGCCCTTCTTCTTGGGACAGTTCCTCAGCTGTCCACC 1587
Db 803 TTGCTGAACACAAACCGTGGCCCTTCTTCTGCTGACAGTTCCTGAGCTGTCCACT 862
QY 1588 TCGCTGCCCTCGCAGTATATACAGGCTCGCCGAGCACTCAGTCCCTCCTGATGCTCAGC 1647
Db 863 TCCCTGCCCGCCGACACATACAGGCTTGGGGAAACCTCAGCCCACTCATGCTCAGC 922
QY 1648 ACCAGTGCACAGGATCACCCCTGTGGAAACCGGACCTTTCGCCCCACGCCCCGCTGCGAAC 1707
Db 923 ACCGACTGGGCTGTATCACCTCTTGGAAACCGGACACACACCCCTCGCAAGTACCAAC 982
QY 1708 CTCTACGGGTCTCACCCCTTCTTCTACCTGGGCTGGAGGACGGCGGTGCGCAACACGGGGTG 1767
Db 983 CTCTACGGGTCTCACCTTCTTCTACCTGGGCTGGAGGACGGTGGCTTGGCTCAGCGTGC 1042
QY 1768 TTCCTGCTTAAACAGCAATGCCATGATGCTGCTGCGAGCCGAGCCCTGCCCTTACGCTGC 1827
Db 1043 TTCCTGCTTAAACAGCAATGCCATGATGCTGCTGCGAAACCCAGCCCACTTAACTGC 1102
QY 1828 AGTGTGACAGGTGGGATCCTGGATGTCTACATCTTCTGGGCGCCAGAGCCCAAGAGCGTG 1887
Db 1103 AGTGTCAACGGGCGGATCCTGGATGTGTATGTTCTTAGGCCAGAGCCCAAGAGCGTC 1162
QY 1888 GTGACAGAGTACTCTGAGCGTTGTGGATACCGTTCATGCTCCGCCATATCTGGGGCTGGGC 1947
Db 1163 GTGCAACAATACCTGATGTTGTGGGATACCCCTTATGCTCCATATCTGGGGCTCTCGGC 1222
QY 1948 TTCACCTGTGCGCTGGGCTACTCTCCACGCTATCACCGCCAGGTGGTGGAGAAC 2007
Db 1223 TTCACCTGTGCGCTGGGCTACTCTCTCGACCGGCAATGTCGCGCAGGTAGTGGAGAAC 1282
QY 2008 ATGACACAGGCGCCACTTCCCTCGTGGAGTCCAAATGGAGAGCTGAGCTACATGAGACTCC 2067
Db 1283 ATGACACAGGACACATCTCCCGTGTGACGCTGCAATGGATGACTGAGCTTACATGAGCGCC 1342
QY 2068 CGGAGGAGTTCACGTTTCAACAAAGATGGCTTCCGGGACTTCCCGGCCATGTTGTCAGGAG 2127
Db 1343 CGAAGAGACTTCACCTTCAACCCAGGACAGCTTTCGCGACTTCCAGACATGTTGGCGGAC 1402
QY 2128 CTGACACAGGCGCGCGGCTTACATGATGATGCTGGATCTCTGCCATCAGCAGCTCGGC 2187
Db 1403 GTGCAACAGGTTGGCGGCTTACATGATGATGCTGGAGCCCTGCCATCAGCAGCGCAGGC 1462
QY 2188 CTGCGCGGAGCTACAGGCCCTTACAGCAGGCTTTCGCGAGGGGCTTTCATCACCACAAAC 2247
Db 1463 CTGCTGGAGGTTACAGGCCCTTACAGCAGGCTTTCGCGAGGGGCGGTTCATCACCACAAAC 1522
QY 2248 GAGACCGGCGCAGCGCTGATTTGGGAAGGTATGGCCCGGGTCCACTGCTTCCCGCACTTC 2307
Db 1523 GAGACTGGGCGCAGCGCTGATTTGGGAAGGTTTCTCTGGAACCAACCGCTTCTCTGATTTTC 1582
QY 2308 ACCAACCCCAAGCGCTTGGCTGTGGGAGGACATGTTGGCTGAGTTCATCAGCAGGCTG 2367
```


[illegible]

Db 504 GCAGGCTCTGCTCAACACAGCGTCTCGCTGTTCTTTGGGAGCAGGTCTCTGCAGAT 563
QY 1581 GTCCACCTCGCTCGCTCGCAGTATATACAGCGCTCGCGAGCAGCTCAGTCCCTGAT 1640
Db 564 CTCACCTCTCGCTCTCGCTTTATTTCTGGCTGGGGAGCGTCTGGCCCTCTCAT 623
QY 1641 GCTCAGCACGAGTGCAGCAGGATCACCTGTGGAAACCGGAGCTTTGCGCCACGCGCG 1700
Db 624 CTTCCGACACGGCTCGACTAAAGTCACGCTCTGGAACCGGACATGGCACCTGCGGCCCA 683
QY 1701 TGGGAACCTCTAGGGTCTCACCTTTCTACCTGGCGCTGGAGGACGGCGGGTGGCACA 1760
Db 684 AGTCACCTCTATGGCTCCACCCCTTCTACCTGGTCTAGAGGATGGTGGCTCTGCACA 743
QY 1761 CGGGGTGTTCTCTAAACAGCAATGCCATGGATGGTCTCGACCGCAGCGCTCCCT 1820
Db 744 TGGTGTCTCTGCTGAAACAGCAATGCCATGGATGGTCTCGACCGCTCGCCAGCCCT 803
QY 1821 TAGCTGGAGGTGCAGAGGTGGATCTGGATGTTACATCTTCTCGGGCCAGAGCCCAA 1880
Db 804 GACCTGGCGCACAAACGGAGGACCTCTGGACTTCTACATCTTCTGGGCCCTGACCCCA 863
QY 1881 GAGCGTGTGCAGCAGTACTGAGACCTTGTGGATACCGCTTCATGCCGCCATACTGGGG 1940
Db 864 GAGCGTGTGCAGCAGTACTGAGATGGTGGTTCCTCTTCATGCCCCCATACTGGGC 923
QY 1941 CTTGGGCTTCCACCTGTGCGCTGGGGCTACTCTCCACGCTATCACCGCCAGGTGT 2000
Db 924 CTTGGGCTTCCACCTGTGCGCTGGGGTATTCTCCACCTCACACCGCGGAGGCTGC 983
QY 2001 GGAGAACATGACAGAGGCCCATCTCCCTCGACGTCCTAATGGAAACGACTTGGACTACAT 2060
Db 984 AGCCAAATGTGAGTGGGCTCTTCCCTCGAGGTGCAATGGAATGACTTGGACTACAT 1043
QY 2061 GGAATCCCGAGGAACTTCACTGTTAAAGAGTGGCTTCGGGACTTCCGGCCATGT 2120
Db 1044 GGAATCCAAAGAGGAACTTCACTTACAAAGGAAACCTTTCAGAGATTAACCTTGACATGT 1103
QY 2121 GCAGGAGTGCACAGGGCGCGCTGCTACATGATGATGATGATGATGATGATGATGATGAT 2180
Db 1104 GCAGGATTCCTATCAGCGTGGCTGCTATACGTCATGATGATGATGATGATGATGATGATGAT 1163
QY 2181 CTCGGGCTTTCGGGAGCTACAGGCTTACAGGCTTACAGGAGGTCTCGGAGGGGGTTCAT 2240
Db 1164 CTCGGGCGCGCGCACCTACCGGCTTATGATGATGATGATGATGATGATGATGATGATGAT 1223
QY 2241 CACCAACGAGACCGGCGCAGCGCTGATGGGAAGGTATGGCCCGGTCCACTGCTTCCC 2300
Db 1224 CCGCAATGCCACGGGGCAGCGCTCATCGGGAAGGTATGGCCAGGCGCCACCGCTTTCCC 1283
QY 2301 CGACTTCACCAACCCACAGCGCTGCTGGTGGGAGGACATGGTGGCTGAGTTCCATGA 2360
Db 1284 AGACTTCACCAACCTTGAGACACAGAGTGGTGGCATGATGATGATGATGATGATGATGATGAT 1343
QY 2361 CCAGGTGCTTTCGAGCGCTTGTGGATGACATGAACGAGCTTCCAACTTTCATCAGAG 2420
Db 1344 GCAGGTGCTTTCGAGCGCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1403
QY 2421 CTCGTAGGACGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2480
Db 1404 CTCACAGGATGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1463
QY 2481 TGGGGGACCTTCAGGCGCCACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2540
Db 1464 TGGGGGCGCTGCTGAGGACAGGACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1523
QY 2541 CTACAACTGCACAACTCTACGGCTGACCGAGGACCATCGCTCCCAAGGGCGCTGT 2600
Db 1524 TTACAACTGCACAGCTGTATGGGCTTAAGTGGGCTTAAGTGGGCTTAAGTGGGCTTAAGTGGG 1583
QY 2601 GAAAGGCTCGGGGACACGCGCAATTTGTGATCTCCGCTCGACTTGTGCTGGCCAGCGCG 2660

Db 1584 CAGGCTGCGGGCACGCGGCCCTTCGTATCTACGCTCCAGCTTTGCGGGACACGGCCA 1643
QY 2661 ATACCGCGCCACTGAGCGGGGACGTGTGGAGCTCTCTGGAGCAGCTCGCTCTCCGT 2720
Db 1644 TTATCAGGGCACTGAGCGGGGATGTGGAGCAGCTGGGAGCAGCTGGCCACTCGT 1703
QY 2721 GCAGAAATCTGAGTTTAACCTGCTGGGGTGTCTGTGGTGGGGCCGACGTCTGCGG 2780
Db 1704 GCAGAGGTGTGCTATTCAACCTCTTGGGTTCCTCTGGTGGGAGCTGACATCTGTGG 1763
QY 2781 CTTCTGCGCAACACCTCTCAGAGAGCTGTGTGCTGCGTGGAGCCAGCTGGGGCCCTTCTA 2840
Db 1764 CTTTGTGTGACACATCTGAGAGCTGTGCTGCTGGAGCCAGCTGGGACCTTCTA 1823
QY 2841 CCCCTTATGCGGAACCAACAGAGCTGTCTAGTCTGCCCCAGGAGCCGTACAGTTCAG 2900
Db 1824 CCCCTTATGAGGAACCAACAGAGCTGTCTAGTCTGCCCCAGGAGCCGTACAGTTCAG 1883
QY 2901 CGAGCGGCGCAGCAGGCGCATGAGGAAGCCCTCACCTGCGCTAGGACTCTCTCCCCA 2960
Db 1884 TCTGCTGACAGGATGCCATGAGGAGTCCCTCGCTCGCTACTCTCTCTCTGCCCCA 1943
QY 2961 CCTCTACACACTGTTCCACAGGCCACGTGCGGGGAGACCGTGGGCCCGCCCTCTT 3020
Db 1944 CTTTACACCTCTTCCACCGGGACACATGCTGGGACACGTTGGGACAGGCGCACTGT 2003
QY 3021 CTTGAGTTTCCCAGAGACTCTAGCACCTTGGACTGTGGACCAACAGCTCTCTGTGGGGGA 3080
Db 2004 CTTGAGTTTCCCAGAGACTCTAGCACCTTGGACTGTGGACCAACAGCTCTCTGTGGGGTGC 2063
QY 3081 GGCCTGTCTATCACCCAGTGTCTCAGGCCGGGAAGGCCGAGTGTGCTGCTTCTTCCC 3140
Db 2064 AGGGCTGTGTATCACACAGTGTCTGGAACAGGAGCAGACAAAGTCAGTGGCTATTTC 2123
QY 3141 CTTGGGCACTGGTAGACCTGTCAGACGTGTCATAGAGGCCCTTTGGCAGCTCCCAAC 3200
Db 2124 AGCTGGACATGGTAGAGCTT----- 2144
QY 3201 CCCACTGCGACTCCCGCTGAGCCAGCCATCCACAGCGGGGAGTGGGTGACGTGCC 3260
Db 2145 -----CACAGGGACTCAACCATCCACAGCAAGGGCAGTGGATCTCTTAGC 2192
QY 3261 GGCCTTGTGACACATCAACGTCCACTCGGGCTGGGTACATCATCCCCCTGAGGG 3320
Db 2193 AGCCCTTGTGACACCAATTAATGTGCATCCGAGCAGGGCACATCTACCCCTGACGA 2252
QY 3321 CCTGGCTCAACCAACAGAGTCCGCGCAGCAGCCCATGGCCCTGGCTGTGGCCCTAAC 3380
Db 2253 GCCTGGATTGAACAGCGGTGAGTCCCGCAAGAGGGGATGATGGTGGTGGCTCTGAC 2312
QY 3381 CAAGGTGGAGAGGCCCGAGGGGAGCTGTTCTGGGACGATGGAGAGAGCTTGAAGTCT 3440
Db 2313 ACCGATGGCTTGGCTGGAGAGCTGTTCTGGGATGATGGGAGAGCTGACAGACTT 2372
QY 3441 GGAGGAGGGGCTTACACAGGTATCTCTGCGCAGGAATAACAGATCGTGAATGA 3500
Db 2373 TGAGAAGGGGACTGCACTGAGATCTCTTCTGCTGCACGCTGGTGGCTGCTCAGTCA 2432
QY 3501 GCTGTGACGTGTGACAGTGGAGGCTGGCTCAGCTGCAAGAGGTGACGTCTCTGGG 3560
Db 2433 GATCTTGGGGCAGGTGGCCACTGGATGGGATCTTGGTGGAGGCTGTCACTGTGCTGG 2492
QY 3561 CTTGCGCAGCGGCGCCAGCAGGTCTCTTCCAAACGCTGTCTCTCTCAAATTTCACCTA 3620
Db 2493 TGTCCCACTGTCTCAGAGGAGTCTTGGCCAAATGGCATCCCGTGGAGGACTTCTCAT 2552
QY 3621 CAGCCCCGACACCAAGGTCTTGAACATCTGTGTCTGCTGTGTTGATGGGAGAGAGTTC 3680
Db 2553 CCGCAGTGACACCCAGGTGTCTCGGTGTCTCGGTGTCTCGCTGCTGCGCCATGTGGGAGCAGTTTGT 2612
QY 3681 CTTGAGCTGGTGTGA 3696
Db 2613 GGTCCCTTGGTCTCTGA 2628

```
RESULT 14
AB000967
LOCUS
DEFINITION
  AB000967 Coturnix japonica GAAI mRNA linear VRT 24-FEB-1999
  cds.
ACCESSION
  AB000967
VERSION
  AB000967.1 GI:3097289
KEYWORDS
  acid alpha glucosidase; GAAI.
SOURCE
  Coturnix japonica (Japanese quail)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
  Phasianinae; Coturnix.
REFERENCE
  1 (sites)
  Kunita,R., Nakabayashi,O., Wu,J.Y., Hagiwara,Y., Mizutani,M.,
  Pennybacker,M., Chen,Y.T. and Kikuchi,T.
  Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail
  (Coturnix coturnix japonica) and the lack of its mRNA in acid
  maltase deficient quails
  Biochim. Biophys. Acta 1362 (2-3), 269-278 (1997)
JOURNAL
  98201735
MEDLINE
  9540858
PUBMED
  9540858
REFERENCE
  2 (bases 1 to 3548)
  Kikuchi,T.
  Direct Submission
  Submitted (10-FEB-1997) Takeki Kikuchi, National Institute of
  Neuroscence, NCNP, Department of Animal Models for Human Disease;
  4-1-1 Ogawahigashimachi, Kodaira, Tokyo 187, Japan
  (E-mail:tkiku@ncnp.nicp.go.jp. Tel:0423-41-2711,
  Fax:0423-46-1754)
FEATURES
  Location/Qualifiers
    1..3548
      /organism="Coturnix japonica"
      /mol_type="mRNA"
      /db_xref="taxon:93934"
      /clone="pQAM8"
      /sex="male"
      /tissue_type="liver"
      /dev_stage="40 day old"
      /note="synonym: Coturnix coturnix japonica"
    1..3548
      /gene="GAAI"
      544..3342
      /gene="GAAI"
      /codon_start=1
      /product="acid alpha glucosidase"
      /protein_id="BAA25884.1"
      /db_xref="GI:3097290"
      /translation="MRSYOKLRATVPOLVLSQEEBEEEDATPTPAGRSKLAPWVVG
      SGLITAVLSTITVWLVQVSGWGHGAPPACQCHLVPSHRYDCYPERNVVVTQELC
      ERGCCFIQTLFVAGKRGVPCFPDPFSPVVSQVSLNQTVMGLLVREKAYPK
      LQMRMDVEFQNTKRLHKITDAANPRYEVELEPRVTVKRAENPIYLSLEISQDPFGV
      LRRQGTQVLLNTVAPLIFADQFLQITLPSFLYGLGHEHSTLHSLDWNLTLL
      WARDAPTESFLNYGAHPFLYLLMEGGDAHGVFLNLSNAMEVALOPAGLTWRTIGGV
      LDFYPLGPDNDYMOQYQEVIFGPAMPPLWALGHLCRWGYSNETWOTAKAMRNF
      QIPQDAQWIDYMGYRDFTPDQKFASLPSLVEDLHKHGOHYVILLDPGISSTSPR
      GAYPPEDEGLRGFLFNTYQQTLLGQVWPQGTATPDFNSNTTHQWLENLORFTHV
      PFDGLWIDNEPSFNMDSGEGPGLSDSPYTPAVLGNLSUTAKTVCASEAQNASVH
      YNLHNLGLKEABEATASALIRGRKPFYISRTPSPQGRYSGHMLGDNRSQMKDYP
      STPGLMSFLGTDILVADICGFSGSTSEELCTRWMLQGFYPPFRNHNKNKAQDP
      TAFSPARTAMKDALTRYSLPLFLVTLFRAHLOQETVARPLFFEPMDVATYGLDR
      QELWGSLIVTPLEPGADSVLGYRPGQVWYDFYTGSSVNSSGEMLKLSAPLDHLNLH
      LREGSLTPQKGIKATRGAPLHLI VALSTRATANGDLFWDDGESLDTFEGQNYSY
      LVFNATENIPTSINVLHSAETADVTIDAVSFYGVQEPSPKVLDDGQKFPFLYDNLQVL
      TVSGGLVLSQGFSLQWL"
      3548
      /gene="GAAI"
      /note="21 A nucleotides"
  polyA_site
    3548
  Query Match
    26.9%; Score 994.4; DB 5; Length 3548;
  ORIGIN
    1121 ACAAGGCCATACCAGGAACAGTGCAGGCGCGCGCTGTCTACATCCCTGCAAGC 1180
    824 ACGTGGTGGTGACCCAGAGCTGTGCGAGAGCGGGGCTGCTTCATCCAGACCCCTCC 883
    1181 AGGGGCTGCAGGAGGCCAGATGGGGCAGCCCTGTGTCTTCTCCACCCAGCTACCCCA 1240
    884 CTGCACT--GGGGGGCAAGCGGGGGTGCCTGTGTCTTACCCACCCGACTTCCCA 940
    1241 GCTACAAGCTGGAGAACCTGAGCTCTCTGAAATGGGCTACACGGCCACCTTGAACCCGTA 1300
    941 GCTACGTGGTGAGAGCTCAACAGACAGTCTGGGCATGACGGGGCTGTGTGTGGCGCA 1000
    1301 CCACCCCCACCTTCTCCCAAGGACATCTGACCTGGGCTGGACGTGATGATGGAGA 1360
    1001 GGGAAAAAGCCCTATTACCCCAAGGACATCCAGATCTGAGGATGGATGGAGTTCACGA 1060
    1361 CTGAGAACCGCTCCACTTTCAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGAT 1420
    1061 CAAACACCGGCTGCACATCAAGATAA CTGATGACAGCAATCCCGCTATGAGGTGCCCC 1120
    1421 TGGAGACCCCGCTGTCCACAGCCGGGCAACCTGCCCACTCTTACAGCGTGGAGTTCCTCG 1480
    1121 TGGAAAGTTCCCGAGTGCACAAAGAGAGACAGAAAAACCCCATCTACAGCCTGGAGATCTCCC 1180
    1481 AGGAGCCCTTCGGGGTGATCGTGCACCGGAGCTGGAGCGCGCTGTCTGTAACACGA 1540
    1181 AGGACCCCTTCGGGGTGATCGTGCACCGGAGCTGGAGCGCGAGTGTCTGCTCAACACA 1240
    1541 CGGTGGCGCCCTGTCTTTCGGGACCAAGTTCCTTTCAGCTGTCCACCTGTCTGCTGCCCTCGC 1600
    1241 CGGTGGCTCCCTGATCTTCGCCGATGATCTTCCTGAGATCTCCACACGCTGCCCTCCA 1300
    1601 AGTATATACAGGCTCTGCGCGAGCACTCTAGTCCCTGATGTCTAGCAGCAGCTGGACCA 1660
    1301 GGTCTCTGTATGGGCTGGGGAGGACCCGAGCACCTGTGACAGCCTCGACTCGAGTGGAA 1360
    1661 GGATCACCTGTGGAAACCGGACCTTGGCGCCACGCGGCTGCGAAGCTCTTACAGGCTCTC 1720
    1361 CACTCACCTCTGGGCGCGTGTATGTGGCGCCCAAGAGTCTTCACTGTACGGGCTC 1420
    1721 ACCCTTCTTACCTGGCGCTGGAGGACGGCGGTGGCGACACGGGGTGTCTCTGTCTAAACA 1780
    1421 ACCCTTCTTACCTGTGATGGAGGAGGCGCGATGTCCTATGGGCTTTTCTCTCTCTCAACA 1480
    1781 GCAATGCCATGGATGTGTCTGCGAGCGGCTGCGCTGCGCTTACCTGTGAGGTGCGAGGTG 1840
    1481 GCAATGCCATGGAGGTGGCGCTGCGAGCCTGCGCGCGGCTGACCTGGAGAACCATTTGGAG 1540
    1841 GGATCCTGTGATGTCTACATCTTCTGGGCGCAGAGCCCAAGAGCGTGGTGCAGCAGTACC 1900
    1541 GTGTCTGCACTTCTACATCTTCTTGGGCGCTGACCCCAACATGATCTCAGCAGTATCC 1600
    1901 TGGAGCTGTGGGATACCCGTTTCATGCCGCCATACTCTGGGCGCTGGGCTTCCACCTGTGCC 1960
    1601 AGGAGTGTATAGTTTCCAGGCAATGCGCCCTCTGTGGCACTTGGCTTCCACCTCTGCC 1660
    1961 GCTGGGGCTACTCTCCACCGCTATCACCCGCGAGGTGGTGGAGAACATGACACCGGGCCC 2020
    1661 GCTGGGGTTACGGCTCCAGCAATGAGACTGGCAGACTGCAAGAGCCCATGAGGAACCTTCC 1720
    2021 ACTTCCCTCTGACCTGCAATGAGAACGACCTGCACTACATGAGACTCCCGAGGGGACTTCA 2080
    1721 AGATCCACAGAGATGCGCAATGGAACGACATCATGATTAATGACCGGGTACCGGGACTTCA 1780
    2081 CGTTCAACAGAGATGCTTCGGGGACTTCCCGGGCCATGTGTGCGAGGTGCAACCGGGGG 2140
    1781 CTTTTGATCCCCAGAAAATTTGGCTCCCTTCACTGTGTGGAAGATCTTCCCAACAGCAG 1840
    2141 GCCGGGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2200
```

```
Db 1841 GGCAGCCTACGTGATAATCTTTGGACCCCGGCATCAGACGACCCACCGTGGCTCT 1900
QY 2201 ACAGGCCCTACGACGAGGCTCTGCGAGGGGGTTCATCACCACGAGACCGGCGAC 2260
Db 1901 ACTGGCTTCGATGAGGCTTGAGACGGGTTTGTTCCTAAACACCCAGGGGCGA 1960
QY 2261 CGCTGATCGGGAAGGATGCGCGGGTCACTGCGCTTCCCGACTTCACAAACCCACAG 2320
Db 1961 CGCTGATCGGCGAGGCTGCGCTGGTTACACTGCTTACCGCTACCCGACTTCTCAAACAGACA 2020
QY 2321 CCCTGGCTGCTGGGAGGACATGGTGGCTGAGTTCCATGACGAGTGCCTTCGAGGCT 2380
Db 2021 CACATCAGTGGTGGCTGGGAACTCGACGGCTTCACACCCAGTACCTTCGATGGCC 2080
QY 2381 TGTGATTCACATGAACGAGGCTTCCAACTTCATCAGAGGCTCTGAGGACGGTGCCTCA 2440
Db 2081 TCTGGATCAGATGATGAGCCATCAAATTCATGGATGGTCTGAAGAGGCTGCCCGC 2140
QY 2441 ACAATGAGCTGAGAAACCAACCTTACGTGCTGGGGTGTGGGGGGAACCTTCCAGGG 2500
Db 2141 CTGGAGAGCTCGACAGCCCGGTACACGCGGCTGTGCTGGGCAATTCCTGACTGCCA 2200
QY 2501 CCACCATCTGTGCTCCAGCCACCTTTCTCTCCACACACTACACCTGCAACCTCT 2560
Db 2201 AGACGGTGTGCTCAGCGGAGCAGACGCAATCGGTGCACTACAACTGCAACCTCT 2260
QY 2561 ACGGCTGACCGAAGCATCGCTCCACAGCGGCGTGGTGAAGGCTCGGGGACACGCG 2620
Db 2261 ACGGCTGAGAGGACCGGAGCGACGCGGAGCGCTTGATCGGATCCGAGGAGCGTC 2320
QY 2621 CATTTGTGATCTCCGCTCGACCTTTGTGGCCAGCGCGGATACGCGGCCACTTGA 2680
Db 2321 CCTTTGTCTATCTCCGCTCCACCTTCCCGAGCCAGGCTCGGTACTCGGGTCACT 2380
QY 2681 GGAAGCTGTGGAGCTCTCGGAGCAGCTCGCTCTCTCGTGCAGAAATCTTGCAGTTTA 2740
Db 2381 GTCAACAAGGAGCAATGAAGGACATGATTAATCTTCACTCCAGGGATGCTGAGCTCA 2440
QY 2741 ACTGCTGGGGTGCCTCTGCTGGGGCGAGCTCTGCGGCTTCTCGGCAACACCTCAG 2800
Db 2441 GCTGTTTGGATCCCACTGTTGGGGGCGGACATCTGTGGCTTCTTGGCAGCAGCTCG 2500
QY 2801 AGGAGCTGTGTGCTGAGACCCAGCTGGGGCTTCTACCCCTTCATGCGGAACCA 2860
Db 2501 AGGAGCTGTGACCCGCTGATGACGCTCGGGCTTCTACCCCTTCTCCCGCAACCA 2560
QY 2861 AAGCTGTCTGCTGCTGCGGAGGCGCTGACGCTTCAAGGAGCGGGCGGAGCAGGCA 2920
Db 2561 ACAACAGAAACGAGAGGCGGAGGACCCACAGCTTTTCAAGCCCTCGGGCAGGACAGCA 2620
QY 2921 TGAGGAAGCCCTCACCTGCGCTAGCGACTCTCTCCCGACCTCTACACACTGTTCCAGC 2980
Db 2621 TGAGGAGCAGCTGCTCACCCGCTACTCTCTCTGCTTCTCTTACAGCTTTTCCATC 2680
QY 2981 AGGCCCACTGCGGGGAGACCGTGGCGCGGCTCTTCTCTGGAGTTTCCCAAGGACT 3040
Db 2681 GTGCCCACTGCAAGGGAGACCGTGGCGCGGCTTGTCTTGTAGTTTCCCTCGGGATG 2740
QY 3041 CTAGCACTTGGACTGTGAGACCAACAGCTCTGTGGGGGAGGCGCTCTCATCACCAG 3100
Db 2741 TGGCCACATACGCGGCTGGACAGGAGTTCCTGTGGGGGAGAGGCTGCTGTTGACACCG 2800
QY 3101 TGCTCAGGCGGAGGAGGAGTCACTGCTGCTTCCCTTGGGCAACATGTTAGAC 3160
Db 2801 TGCTGAGCGCGGAGCAGACTCAGTGTGGGCTATTTCCCGCAGGGGCTGTGTGAGT 2860
QY 3161 TGCAAGCGGTGCCAATAGAGGCGCTTGGCAGCCTCCACCCCGACCTGAGCTCCCGGTG 3220
Db 2861 TCTACAG-----G 2869
QY 3221 AGCCAGCCATCCACAGCGAGGCGAGTGGGTGACGCTGCGGCGCCCTCGGACCACTCA 3280
Db 2870 GCTCCTCTGTGAACAGCAGTGGGGAGATGTTAAAGCTGTAAAGCTGTGAGCCCGCTG 2929
```

```
QY 3281 AGTCCACCTCGGGCTGGGTACATCATCCCTCGAGGGCCCTGGCTCTCAACACACAG 3340
Db 2930 ACCTGCATCTCGGAGGGTTCATCTCTGCCACACAGAAACCGGGATCACAGCAAG 2989
QY 3341 AGTCCCGCAGCAGCCATGCGCTGGCTGTGGCCCTAACCAAGGTGGAGAGGCCGAG 3400
Db 2990 CCACCGAGGGAACCTCTGCATCTCATTTGGGCCCTGTCCACACGTGCCACTGCTGG 3049
QY 3401 GGGAGCTGTTCTGGGACGATGGAGAGAGCTGGAAAGTGTGGAGGAGGGGCTTACAC 3460
Db 3050 GGGACCTTTCTGGGATGATGGAGAGAGCTGGACACCTTCGAGCAGGGAAGTACTCT 3109
QY 3461 AGGTCACTTCTGCGCCAGGAATAACAGATCGTGAATGAGCTGGTACGTGTGACAGTG 3520
Db 3110 ACCTGGTGTCAATGCCACGGAGAACATCTTCACCTCCAAACGTCTTATGCGCAGCAC 3169
QY 3521 AGGAGCTGGCTGCGAGCTGCGAGAGTGTCTCTGGGCTGCGCAGGCGGCCCGCAG 3580
Db 3170 AGGCCACCGATGTACCATCGATGAGTGTCTATGAGTGCAGGAGCGGCCGAGCA 3229
QY 3581 AGGTCTCTCCAAACGG 3596
Db 3230 AAGTCTCTCGATGG 3245

RESULT 15
CQ776406
LOCUS CQ776406 1493 bp DNA linear PAT 11-MAR-2004
DEFINITION Sequence 92 from Patent EP1394274.
ACCESSION CQ776406
VERSION CQ776406.1 GI:45379796
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ohnani,N., Sugita,Y., Yamaya,M., Kubo,H., Nagai,H. and Izuwara,K.
TITLE Methods of testing for bronchial asthma or chronic obstructive
JOURNAL Patent: EP 1394274-A 92 03-MAR-2004;
FEATURES
source Genex Research, Inc. (JP)
location/Qualifiers
1..1493
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 25.9%; Score 957; DB 6; Length 1493;
Best Local Similarity 99.0%; Pred. No. 1.5e-138;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 63 CGGATCTTACTCGGGGAGAGAAACAGCCCAAGCGCTCCCGAAACGGAGTCCGGAGA 122
Db 109 CGGCAAGTACTCGGGGAGAGAAACAGCCCAAGCGCTCCCGAAACGGAGTCCGGAGA 168
QY 123 GGAGTTCGCGATGGAGAGTTGAACCGCTGTGGGAGAGGCCCGCAGCTGCATCTTCC 182
Db 169 GGAGTTCGCGATGGAGAGTTGAACCGCTGTGGGAGAGGCCCGCAGCTGCATCTTCC 228
QY 183 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 242
Db 229 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 288
QY 243 CTGGAAGAACTAAAGCTTGACGGCTTGAACGAGATGGGGAGAGAGAGCGAGACTCAT 302
Db 289 CTGGAAGAACTAAAGCTTGACGGCTTGAACGAGATGGGGAGAGAGAGCGAGACTCAT 348
QY 303 AGCAACCTCAATGATCATCTTGGCCCAAGTATGGTCTGGACGGAAGAGAGCGCTCGCA 362
Db 349 AGCAACCTCAATGATCATCTTGGCCCAAGTATGGTCTGGACGGAAGAGAGCGCTCGCA 408
```


QY	363	GGTGACCAAGCACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGGA	422
Db	409	GGTGACCAAGCACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGGA	468
QY	423	AAAGCTGTGGCACAAGCGCAAGACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT	482
Db	469	AAAGCTGTGGCACAAGCGCAAGACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAGCT	528
QY	483	CTGGCGGGAGTTCTCTCATCACAAAGAGAAAGTTACAGAGTACAACTCTCTGTGGAGAC	542
Db	529	CTGGCGGGAGTTCTCTCATCACAAAGAGAAAGTTACAGAGTACAACTCTCTGTGGAGAC	588
QY	543	CCTGAGCAGGACCGAAGAAATCCACGAGAACGTCATTAGCCCTCGGACCTTGAGCGACAT	602
Db	589	CCTGAGCAGGACCGAAGAAATCCACGAGAACGTCATTAGCCCTCGGACCTTGAGCGACAT	648
QY	603	CAAGGCGAGGCTCTGCACAGCAGGACACACGAGCTGAAGGAGAGCTCGGACGATCAA	662
Db	649	CAAGGCGAGGCTCTGCACAGCAGGACACACGAGCTGAAGGAGAGCTCGGACGATCAA	708
QY	663	CCAGGCGCTGGACCCGCTCGCAGGCTCAGCCACAGGGCTACAGCACTGAGGCTGAGTT	722
Db	709	CCAGGCGCTGGACCCGCTCGCAGGCTCAGCCACAGGGCTACAGCACTGAGGCTGAGTT	768
QY	723	CGAGGAGCCCAAGGGTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACTCACGACAA	782
Db	769	CGAGGAGCCCAAGGGTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACTCACGACAA	828
QY	783	GGAGCTGGAGGGCTTCGGGAGGAGCTCAAGCACTTCGAAGCCMAATCGAGAAGCACAA	842
Db	829	GGAGCTGGAGGGCTTCGGGAGGAGCTCAAGCACTTCGAAGCCMAATCGAGAAGCACAA	888
QY	843	CCACTACCAAGACAGCTGGAGATTCCGCACAGAGAGCTGAGGCACGACAGAGCGTGGG	902
Db	889	CCACTACCAAGACAGCTGGAGATTCCGCACAGAGAGCTGAGGCACGACAGAGCGTGGG	948
QY	903	CGACGGCGAGCGTGTGAGCCGCGACCGCGAGAGCAACGCCCTCTGTGGAGGGCGGACCAA	962
Db	949	CGACGGCGAGCGTGTGAGCCGCGACCGCGAGAGCAACGCCCTCTGTGGAGGGCGGACCAA	1008
QY	963	GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC	1022
Db	1009	GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC	1068
QY	1023	TCGCGCCGAGGCA	1035
Db	1069	TCGGCACACGAA	1081

Search completed: April 4, 2005, 15:52:18
Job time : 15278 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2005, 11:07:54 ; Search time 1756 Seconds
(without alignments)
12480.001 Million cell updates/sec

Title: US-10-600-862A-6
Perfect score: 3702
Sequence: 1 cttaccgccatcggggtcc.....tcagctggtgactcgag 3702

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2647.8	71.5	3026	13 ADQ91897	Adq91897 Human glu
2	2647.8	71.5	3624	10 ACf80584	Acf80584 Human alp
3	2647.8	71.5	3846	13 ADQ91895	Adq91895 Human glu
4	2636.6	71.2	3847	13 ACn40294	Acn40294 Tumour-as
5	2635.4	71.2	2851	12 ADf47520	Adf47520 Human GNA
6	2186.2	59.1	3376	13 ACn42967	Acn42967 Human dia
7	2097.2	56.7	3455	13 ACn42966	Acn42966 Human dia
8	1887.4	51.0	3105	13 ACn42968	Acn42968 Human dia
9	1881.4	50.8	3364	6 ABi99468	Abi99468 Mouse isc
10	1191.6	32.2	2655	13 ACN42969	ACN42969 Human dia
11	957	25.9	1493	5 AAC87527	Aac87527 Human RAP
12	957	25.9	1493	6 AAL48893	Aal48893 Human rec
13	957	25.9	1493	6 ABK84062	Abk84062 Human cdn
14	957	25.9	1493	12 ADJ74840	Adj74840 Marker ge
15	957	25.9	1493	13 ADf24896	Adf24896 Breast ca
16	957	25.9	1493	13 ADP23134	Adp23134 PRO:polyd
17	954.2	25.8	2084	13 ADQ38777	Adq38777 Human SNP
18	954.2	25.8	2106	13 ADQ38776	Adq38776 Human SNP
19	953.8	25.8	1499	2 AAz33614	Aaz33614 Human bre
20	665.6	18.0	2902	6 ABi99326	Abi99326 Mouse isc

c	21	665.6	18.0	2902	12	ADJ75755	Adj75755 Marker ge
	22	437.2	11.8	470	9	ACH44040	Ach44040 Human foe
	23	411	11.1	956	6	ABN74475	Abn74475 Bovine em
	24	372.4	10.1	421	6	ABV94943	Abv94943 Human pan
	25	371.2	10.0	5825	9	ACH03918	Ach03918 Human cdn
	26	371.2	10.0	6483	4	AAH57440	Aah57440 Human int
	27	371.2	10.0	6513	12	ADO19559	Ado19559 Human PRO
	28	370.6	10.0	449	6	ABT07013	Abt07013 Human ova
	29	370.6	10.0	449	8	ABX72891	Abx72891 Human ova
	30	369.8	10.0	2752	2	AAV11736	Aav11736 Barley al
	31	368	9.9	1644	10	ADB53939	Adb53939 Primary r
	32	368	9.9	1644	10	ABT42500	Abt42500 Toxicity
	33	368	9.9	1644	12	ADP73017	Adp73017 Renal tox
	34	361	9.8	991	6	ABN74476	Abn74476 Bovine em
	35	338.8	9.2	18272	4	AAK69447	Aak69447 Human imm
	36	338.8	9.2	18272	4	AAU06181	Aal06181 Human rep
	37	338.8	9.2	18272	4	ABL98746	Abi98746 Human tea
	38	318.6	8.6	466	9	ACH48464	Ach48464 Human leu
	39	298.2	8.1	465	9	ACH18185	Ach18185 Human adu
	40	284	7.7	649	13	ADQ52009	Adq52009 Novel can
c	41	284	7.7	667	10	ADB57541	Adb57541 Toxicity-
	42	284	7.7	667	10	ADB52060	Adb52060 Primary r
c	43	284	7.7	667	12	ADP72364	Adp72364 Renal tox
	44	279.2	7.5	1560	10	ADC07849	Adc07849 Rice DNA
	45	278.6	7.5	1560	10	ADC08230	Adc08230 Rice DNA

ALIGNMENTS

RESULT 1
ADQ91897
ID ADQ91897 standard; cdna; 3026 BP.

XX ADQ91897;

XX 21-OCT-2004 (first entry)

XX Human glucosidase alpha acid (GAA) encoding cdna SEQ ID NO:3.

XX chimeric protein; secretory signal; lysosomal protein;

KW lysosomal acid alpha-glucosidase; glucosidase alpha acid; GAA; enzyme;

KW lysosomal acid alpha-glucosidase deficiency; GAA deficiency;

KW glycogen storage disease type II; GSD II; human; chromosome 17; gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 33..2891

FT /*tag= a

FT /product= "glucosidase alpha acid (GAA) "

XX WO2004064750-A2.

XX 05-AUG-2004.

XX 21-JAN-2004; 2004WO-US001453.

XX 22-JAN-2003; 2003US-0441789P.

XX (UYDU-) UNIV DUKE.

XX Koeberl DD, Sun B;

XX WPI; 2004-571599/55.

XX P-PSDB; ADQ91896.

XX GENBANK; NM_000152.

XX New nucleic acid expressing lysosomal acid alpha-glucosidase (GAA)

PT polypeptide, useful for preparing a composition for treating e.g.,

XX glycogen storage disease type II.

PS Claim 45; SEQ ID NO 3; 128pp; English.

The present invention describes an isolated nucleic acid (I) encoding a chimeric polypeptide comprising a secretory signal sequence operably linked to a lysosomal polypeptide. Also described: (1) a vector comprising the isolated nucleic acid; (2) a pharmaceutical formulation comprising the isolated nucleic acid in a carrier; (3) a cell comprising the isolated nucleic acid; (4) a chimeric polypeptide comprising a secretory signal sequence operably linked to a lysosomal polypeptide; (5) delivering a nucleic acid encoding a lysosomal polypeptide (preferably a lysosomal acid alpha-glucosidase (GAA)) to a cell; (6) producing a GAA polypeptide in a cultured cell; (7) treating a deficiency of a lysosomal polypeptide or lysosomal acid alpha-glucosidase (GAA) in a subject; and (8) an isolated nucleic acid (II) encoding a GAA polypeptide comprising: (a) a coding region encoding a GAA polypeptide; and (b) a 3' untranslated region where: (i) the 3' untranslated region (UTR) comprises a GAA 3' UTR comprising a deletion of at least 25 consecutive nucleotides, so that upon introduction into a cell, GAA polypeptide is produced at a higher level from the nucleic acid as compared with its production from a nucleic acid comprising a full-length GAA 3' UTR; or (ii) the 3' UTR is less than 200 nucleotides in length and comprises a segment that is heterologous to the GAA coding region, so that GAA polypeptide is overexpressed on introduction of the nucleic acid into a cell. (I) has hepatotropic activity, and can be used in gene therapy. The nucleic acid (I) encoding a chimeric polypeptide comprising a secretory signal sequence operably linked to a lysosomal acid alpha-glucosidase (GAA) polypeptide is useful for preparing a composition for treating lysosomal acid alpha-glucosidase (GAA) deficiency e.g., glycogen storage disease type II (GSD II). The present sequence encodes the human GAA amino acid sequence, which is used in the exemplification of the present invention. The human GAA gene is located on chromosome 17, more specifically to 17q25.2-q25.3.

SQ Sequence 3026 BP; 544 A; 1040 C; 901 G; 541 T; 0 U; 0 Other;

Query Match	71.5%;	Score 2647.8;	DB 13;	Length 3026;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 2652;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
QY	1036	GAACCGGTGCACACCCCGCGCTCCAGAGCAGTGCACACAGTGCAGCTCCCCCC	1095	
DB	231	GATGCCCAGGACACCCCGCGCTCCAGAGCAGTGCACACAGTGCAGCTCCCCCC	290	
QY	1096	AACAGCCGCTTCGATTGCGCCCTTGACAAGGCCCATCAACCAGGAACAGTGCAGGCCCGC	1155	
DB	291	AACAGCCGCTTCGATTGCGCCCTTGACAAGGCCCATCAACCAGGAACAGTGCAGGCCCGC	350	
QY	1156	GGCTGCTGTACATCCCTGTCAAGACAGGGGCTGCAGGAGGCCACAGATGGGGCAGCCCTGG	1215	
DB	351	GGCTGCTGTACATCCCTGTCAAGACAGGGGCTGCAGGAGGCCACAGATGGGGCAGCCCTGG	410	
QY	1216	TGCTTCTTCCACCCACGCTACCCAGCTACAAGCTGGAGAACCTTGAGCTCTCTTGAATG	1275	
DB	411	TGCTTCTTCCACCCACGCTACCCAGCTACAAGCTGGAGAACCTTGAGCTCTCTTGAATG	470	
QY	1276	GGCTACACGGCCACCTTGACCCGTACCACCCCACTTCTTCCCAAGGACATCTCTGACC	1335	
DB	471	GGCTACACGGCCACCTTGACCCGTACCACCCCACTTCTTCCCAAGGACATCTCTGACC	530	
QY	1336	CTGGGGCTGACGTGATGATGGAGACTGAGAACCGCCCTCACTTCAAGATCAAGATCCA	1395	
DB	531	CTGGGGCTGACGTGATGATGGAGACTGAGAACCGCCCTCACTTCAAGATCAAGATCCA	590	
QY	1396	GCTAACAGGCGCTACGAGGTGCCCTTGGAGACCCCGGTGTCCACACCGGGGACCGTCC	1455	
DB	591	GCTAACAGGCGCTACGAGGTGCCCTTGGAGACCCCGGTGTCCACACCGGGGACCGTCC	650	
QY	1456	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGCACCCGGCAGCTG	1515	
DB	651	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGCACCCGGCAGCTG	710	
QY	1516	GACGGCCGCTGTCTGTGAACAGACGGTGGCGCCCTGTCTTTTTCGGGACCAAGTTCCTT	1575	
DB	711	GACGGCCGCTGTCTGTGAACAGACGGTGGCGCCCTGTCTTTTTCGGGACCAAGTTCCTT	770	

Qy	2656	GGCCGATACCGCGGCCACTGACACGGGGACGCTGTGGAGCTCTCTGGAGACAGCTCGCCCTCC	2711
Db	1851	GGCCGATACGCGCGGCACCTGACACGGGGACGCTGTGGAGCTCTCTGGAGACAGCTCGCCCTCC	1910
Qy	2716	TCCGTGCCAGAAATCTCTGCAGTTTAACTCTGCTGGGGGTGCCCTCTGTGTCGGGGCCGACGTC	2775
Db	1911	TCCGTGCCAGAAATCCTGCACTTTAACTCTGCTGGGGGTGCCCTCTGTGTCGGGGCCGACGTC	1970
Qy	2776	TGCGGCTTCCTGGGCAACACCTCAGAGAGCTGTGTGTCGCTGAGACCCAGACTGGGGGCC	2835
Db	1971	TGCGGCTTCCTGGGCAACACCTCAGAGAGCTGTGTGTCGCTGAGACCCAGACTGGGGGCC	2030
Qy	2836	TTCTACCCCTTCATGCGGAACCAACAAGACCTGCTCAGTCTGCGCCCAAGGAGCCGTACAGC	2895
Db	2031	TTCTACCCCTTCATGCGGAACCAACAAGACCTGCTCAGTCTGCGCCCAAGGAGCCGTACAGC	2090
Qy	2896	TTCAGCGAGCCGGCCACAGCAGGCCATGAGGAAGCCCTCAACCCTGCCTAAGCACTCCTC	2955
Db	2091	TTCAGCGAGCCGGCCACAGCAGGCCATGAGGAAGCCCTCAACCCTGCCTAAGCACTCCTC	2150
Qy	2956	CCCCACCTCTACACACTGTGTTCCACACAGGCCACAGCTCGCGGGGAGACGCTGGCCCGGCC	3015
Db	2151	CCCCACCTCTACACACTGTGTTCCACAGGCCACAGCTCGCGGGGAGACGCTGGCCCGGCC	2210
Qy	3016	CTCTTCTTGGAGTTCCTCCAAAGCACTCTAGCACCTTGGACTGTGACACCAACAGCTCTCTGG	3075
Db	2211	CTCTTCTTGGAGTTCCTCCAAAGCACTCTAGCACCTTGGACTGTGACACCAACAGCTCTCTGG	2270
Qy	3076	GGGAGAGCCCTGCTCATCAACCCAGTGTGTCAGGCCGGGAGGCGCAAGTGACTGGGTAC	3135
Db	2271	GGGAGAGCCCTGCTCATCAACCCAGTGTGTCAGGCCGGGAGGCGCAAGTGACTGGGTAC	2330
Qy	3136	TTCCCTTTGGSCACATGATAGCACTCAGACGCTGCAATAGAGGCCCTTGGCAGCCTC	3195
Db	2331	TTCCCTTTGGSCACATGATAGCACTCAGACGCTGCAATAGAGGCCCTTGGCAGCCTC	2390
Qy	3196	CCACCCCACTGACGTCCCGCTGAGCCAGCCATCCACAGCAGGGGGCAGTGGGTGACG	3255
Db	2391	CCACCCCACTGACGTCCCGCTGAGCCAGCCATCCACAGCAGGGGGCAGTGGGTGACG	2450
Qy	3256	CTGCGCGCCCTTGAGACACCATCAACGTCCACCTCGGGCTGGGTACATCATCCCCCTG	3315
Db	2451	CTGCGCGCCCTTGAGACACCATCAACGTCCACCTCGGGCTGGGTACATCATCCCCCTG	2510
Qy	3316	CAGGGCCCTTGGCTCACAACACAGAGTCCCGCCAGCAGCCCATGGGCCCTGGGTGTGGCC	3375
Db	2511	CAGGGCCCTTGGCTCACAACACAGAGTCCCGCCAGCAGCCCATGGGCCCTGGGTGTGGCC	2570
Qy	3376	CTAACCAAGGGTGAGAGAGCCGAGGGGAGCTGTTCTTGGACGATGAGAGAGCCCTGGAA	3435
Db	2571	CTAACCAAGGGTGAGAGAGCCGAGGGGAGCTGTTCTTGGACGATGAGAGAGCCCTGGAA	2630
Qy	3436	GTGCTGAGCGAGGGGCTTACACACAGTCACTTCTCTGCGCCAGGAATAACAGATCGTG	3495
Db	2631	GTGCTGAGCGAGGGGCTTACACACAGTCACTTCTCTGCGCCAGGAATAACAGATCGTG	2690
Qy	3496	AATGAGCTGTGACGTGTGACAGTGAAGGAGCTGGCTGACGTGCAAGAGGTGACTGTG	3555
Db	2691	AATGAGCTGTGACGTGTGACAGTGAAGGAGCTGGCTGACGTGCAAGAGGTGACTGTG	2750
Qy	3556	CTGGGCTGGCCACGGCGCCACAGAGGTCTCTCCAAACGGTTCCTGTCTCCAACTTC	3615
Db	2751	CTGGGCTGGCCACGGCGCCACAGAGGTCTCTCCAAACGGTTCCTGTCTCCAACTTC	2810
Qy	3616	ACCTACAGCCCGCACACAAAGTCTCTGGACATCTGTGCTCTCGCTGTTGATGGGAGAGCAG	3675
Db	2811	ACCTACAGCCCGCACACAAAGTCTCTGGACATCTGTGCTCTCGCTGTTGATGGGAGAGCAG	2870
Qy	3676	TTTCTCTGCTGAGCTGGTGT 3694	
Db	2871	TTTCTCTGCTGAGCTGGTGT 2889	

RESULT 2	
ACF80584	standard; cDNA; 3624 BP.
XX	
AC	ACF80584;
XX	
DT	15-JAN-2004 (first entry)
XX	
DE	Human alpha-glucosidase coding sequence.
XX	
KW	Human; alpha-glucosidase; lysosome; enzyme; Pompe disease;
KW	transgenic plant; gene; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	220..3078
FT	/*tag= a
FT	/product= "Human alpha-glucosidase"
FT	220..426
FT	/*tag= b
FT	427..3075
FT	/*tag= c
XX	
EN	WO2003073839-A2.
XX	
PD	12-SEP-2003.
XX	
PF	03-MAR-2003; 2003WO-IT000120.
XX	
PR	01-MAR-2002; 2002IT-RM000115.
XX	
PA	(PLAN-) PLANTECHNO SRL.
XX	
PI	Fogher C, Reggi S;
XX	
WI	WPI: 2003-712829/67.
DR	P-PSDB; ABM79002.
DR	
XX	
PT	New genetically transformed plant that can produce a lysosomal
PT	animal or human origin, useful for preparing a medicament for
PT	replacement therapy in Gaucher, Anderson-Fabry or Pompe disease
XX	
PS	Example 10; Page 49-51; 53pp; English.
XX	
CC	The present sequence is the coding sequence for human alpha-
CC	(GAA) . A deficiency of this enzyme causes Pompe disease. The
CC	based on the discovery that lysosomal enzymes, such as GAA
CC	expressed in seed storage organs in a form which is stable
CC	months in stored seeds), enzymatically active and in a high
CC	suitable for medical use. GAA cDNA was obtained from placental
CC	PCR using primers that deleted the signal sequence and nat
CC	It was used to construct vector pPLT4200, which also includ
CC	7S soy globulin promoter (PGL0B, see ACF80579) and the bas
CC	globulin signal sequence (see ACF80580) upstream of the ve
CC	site. Such constructs can be used to express human GAA in t
CC	plants, especially in plants having a high protein content
CC	cereals and tobacco
XX	
SQ	Sequence 3624 BP; 643 A; 1220 C; 1100 G; 661 T; 0 U; 0 Other
	Query Match 71.5%; Score 2647.8; DB 10; Length
	Best Local Similarity 99.7%; Pred. No. 0;
	Matches 2652; Conservative 0; Mismatches 7; Indels
QY	1036 GAACACCGGTGCACACCCCGCGCGTCCACAGAGCAGTGCACACAGTGC
Db	418 GATGCCCAGGCACACCCCGCGCGTCCAGAGCAGTGCACACAGTGC
QY	1096 AACAGCGCTTCGATTGGCCCGCTTACAAAGGCCATCACCAGGAACAGT
Db	478 AACAGCGCTTCGATTGGCCCGCTTACAAAGGCCATCACCAGGAACAGT

QY	1156	GGCTGCTGCTACATCCCTGCAAGCAGGGGCTGACAGAGCCAGATGGGGCAGCCCTGG	1215		Db	1618	TTTCATCAACAACGACCGGGCAGCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCC	1677	
Db	538	GGCTGCTGCTACATCCCTGCAAGCAGGGGCTGACAGAGCCAGATGGGGCAGCCCTGG	597		QY	2296	TTTCCCGGACTTACAAACCCACACGCGCTGGCTGGTGGAGGACATGGTGGCTGAGTTC	2355	
QY	1216	TGCTTTCTCCACCCAGCTTACCCAGCTACAGCTGGAGAACCTGAGCTCTCTGAAATG	1275		Db	1678	TTTCCCGGACTTACAAACCCACACGCGCTGGCTGGTGGAGGACATGGTGGCTGAGTTC	1737	
Db	598	TGCTTTCTCCACCCAGCTTACCCAGCTACAGCTGGAGAACCTGAGCTCTCTGAAATG	657		QY	2356	CATGACCAAGGTGCCCTTCGACGGCTTGTGGATTGACATGAACGAGCCTTCCAATTTCACTC	2415	
QY	1276	GGCTACACGGCCACCTGACCGGTACACCCCCACCTTCTTCCCAGGACATCTGTACC	1335		Db	1738	CATGACCAAGGTGCCCTTCGACGGCATGTGATTTGACATGAAGAGCCTTCCAATTTCACTC	1797	
Db	658	GGCTACACGGCCACCTGACCGGTACACCCCCACCTTCTTCCCAGGACATCTGTACC	717		QY	2416	AGAGGCTCTGAGGACGGCTGCCCCAAATGAGCTGGAGAACCCACCTTACCTGCTGGG	2475	
QY	1336	CTGCGGCTGAGAGTGATGAGAGCTGAGAACCGGCTTCCATTTCAAGTCAAAAGATCCA	1395		Db	1798	AGAGGCTCTGAGGACGGCTGCCCCAAATGAGCTGGAGAACCCACCTTACCTGCTGGG	1857	
Db	718	CTGCGGCTGAGAGTGATGAGAGCTGAGAACCGGCTTCCATTTCAAGTCAAAAGATCCA	777		QY	2476	GTGGTTGGGGGAGCCCTCCAGCGGCCCACTCTGTGCTCCAGCACACAGTTTCTCTCC	2535	
QY	1396	GCTAACAGGGGCTACAGAGTGCCCTTGGAGACCCCGCGTGTCCACAGCGGGCACCGTCC	1455		Db	1858	GTGGTTGGGGGAGCCCTCCAGCGGCCCACTCTGTGCTCCAGCACACAGTTTCTCTCC	1917	
Db	778	GCTAACAGGGGCTACAGAGGTGCCCTTGGAGACCCCGCGTGTCCACAGCGGGCACCGTCC	837		QY	2536	ACACACTACAACTTCACAAACCTCTACCGGCTTGACCGAAGCCATCGCCTTCCACAGGGCG	2595	
QY	1456	CCACTCTACAGCGTGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGACACGGGAGCTG	1515		Db	1918	ACACACTACAACTTCACAAACCTCTACCGGCTTGACCGAAGCCATCGCCTTCCACAGGGCG	1977	
Db	838	CCACTCTACAGCGTGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGACACGGGAGCTG	897		QY	2596	CTGGTGAAGGCTCGGGGAGACGCCCATTTGTGATCTCCCGCTCGACCTTTCTGCTGGCCAC	2655	
QY	1516	GACGGCCGGTGCTGTGAACACGACGGGTGGGCCCTCTGTCTTTTTCGGACACAGTTTCCTT	1575		Db	1978	CTGGTGAAGGCTCGGGGAGACGCCCATTTGTGATCTCCCGCTCGACCTTTTCTGCTGGCCAC	2037	
Db	898	GACGGCCGGTGCTGTGAACACGACGGGTGGGCCCTCTGTCTTTTTCGGACACAGTTTCCTT	957		QY	2656	GGCCGATACGGCGGCACTTGGACGGGGAGCGTGTGGAGCTCTCTGGAGAGCAGCTCCCTCC	2715	
QY	1576	CAGCTGTCAACCTGCTGCTCGAGTATATACAGAGGCTTGCCTGAGACACCTCAGTCC	1635		Db	2038	GGCCGATACGGCGGCACTTGGACGGGGAGCGTGTGGAGCTCTCTGGAGAGCAGCTCCCTCC	2097	
Db	958	CAGCTGTCAACCTGCTGCTCGAGTATATACAGAGGCTTGCCTGAGACACCTCAGTCC	1017		QY	2716	TCGGTCCAGAAATCCTTCAGGTTTAACTGTCTGGGGGTGCTCTGCTGGGGCCGACGCTC	2775	
QY	1636	CTGATGCTCAGCACAGCTGGACCAAGGATCACTCTGTGAACCGGGACCTTGGCCCCAG	1695		Db	2098	TCCGTGCCAGAAATCCTTCGAGTTTAACTGTCTGGGGGTGCTCTGCTGGGGCCGACGCTC	2157	
Db	1018	CTGATGCTCAGCACAGCTGGACCAAGATCACTCTGTGAACCGGGACCTTGGCCCCAG	1077		QY	2776	TGCGGCTTCTGGGGCAACACTCAGAGGAGCTGTGTGGCTGAGACCCAGCTGGGGGCC	2835	
QY	1696	CCGGTTCGAAACCTCTACCGGCTCAACCTTTCTACTCGCGCTGGAGACCGCGGGTGC	1755		Db	2158	TGCGGCTTCTGGGGCAACACTCAGAGGAGCTGTGTGGCTGAGACCCAGCTGGGGGCC	2217	
Db	1078	CCGGTTCGAAACCTCTACCGGCTCAACCTTTCTACTCGCGCTGGAGACCGGGTGC	1137		QY	2836	TTCTACCCCTTCATCGGGAAACCAACAGCCTGCTCAGTCTGCCCCAGAGCCGTACACAG	2895	
QY	1756	GCACAGGGGTGTTCTCTGCTAAACAGCAATGCCATGGATGTGGTCTGACGGGAGCCCT	1815		Db	2218	TTCTACCCCTTCATCGGGAAACCAACAGCCTGCTCAGTCTGCCCCAGAGCCGTACAGC	2277	
Db	1138	GCACAGGGGTGTTCTCTGCTAAACAGCAATGCCATGGATGTGGTCTGACGGCGAGCCCT	1197		QY	2896	TTACGCGAGCCGGCCAGCAGCCCATGAGGAGGCCCTCACCTCGCTGCTACGCACTCTCTC	2955	
QY	1816	GCCTTTAGCTGGAGGTGCAAGAGTGGATCTCTGATGTCTACATCTTCTGGGCCAGAG	1875		Db	2278	TTACGCGAGCCGGCCAGCAGGCCCATGAGGAGGCCCTCACCTCGCTGCTACGCACTCTCTC	2337	
Db	1198	GCCTTTAGCTGGAGGTGCAAGAGTGGATCTCTGGATGTCTACATCTTCTGGGCCAGAG	1257		QY	2956	CCCGACCTTACACACTGTTCCACAGGCCACGTCGGGGGGAGACCGTGGCCCGGGCC	3015	
QY	1876	CCCAAGAGCGTGTGACAGTACCTTGGACGTTGTGGGATACCCGTTCAATGCGGCCATAC	1935		Db	2338	CCCGACCTTACACACTGTTCCACAGGCCCACTGTCGGGGGGAGACCGTGGCCCGGGCC	2397	
Db	1258	CCCAAGAGCGTGTGACAGTACCTTGGACGTTGTGGGATACCCGTTCAATGCGGCCATAC	1317		QY	3016	CTCTTCTTGGAGTTCCTCCAGAGACTCTAGCACTGGACTGTGGAGCACACAGCTCTGTGG	3075	
QY	1936	TGGGGCTTGGGCTTCCACCTGTCGCGCTGGGGTACTCTCTCCACCGCTATACCCCGCCAG	1995		Db	2398	CTCTTCTTGGAGTTCCTCCAGAGACTCTAGCACTGGACTGTGGAGCACACAGGCTCTGTGG	2457	
Db	1318	TGGGGCTTGGGCTTCCACCTGTCGCGCTGGGGTACTCTCTCCACCGCTATACCCCGCCAG	1377		QY	3076	GGGGAGGCCCTGCTCATCACTTCCAGGCTGCTCAGGCGGGGAGCGGAGTGAAGTGAAGTAC	3135	
QY	1996	GTGGTGGAGAACATGACAGGGGCCACTTCCCGCTGGACGTCCTCAATGGAAACGACCTGGAC	2055		Db	2458	GGGGAGGCCCTGCTCATCACTTCCAGGCTGCTCCAGGCGGGAGCGGAGGCTGAGTGAAGTAC	2517	
Db	1378	GTGGTGGAGAACATGACAGGGGCCACTTCCCGCTGGACGTCCTCAATGGAAACGACCTGGAC	1437		QY	3136	TTCCCTTGGGACATGGTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC	3195	
QY	2056	TACATGGACTCCCGAGGAGCTTTCAGCTTCAACAGAGATGGCTTCGCGGACTTCCCGGCC	2115		Db	2518	TTCCCTTGGGACATGGTACGACCTGACAGCGGTGCCAATAGAGGCCCTTGGCAGCCTC	2577	
Db	1438	TACATGGACTCCCGAGGAGCTTTCAGCTTCAACAGAGATGGCTTCGCGGACTTCCCGGCC	1497		QY	3196	CCACCCCACTGACAGCTCCCGTGGAGCCATCCACAGGAGGGGAGTGGGTGAGC	3255	
QY	2116	ATGGTGCAGAGCTGACACAGGGCGCGCGCTACATGATGATCTGTGATCTGTGCATC	2175		Db	2578	CCACCCCACTGACAGCTCCCGTGGAGCGGCCATCCACAGGAGGGGAGTGGGTGAGC	2637	
Db	1498	ATGGTGCAGAGCTGACACAGGGCGCGCGCTACATGATGATCTGTGATCTGTGCATC	1557		QY	3256	CTGCGCGGCCCTTGGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCTG	3315	
QY	2176	AGCAGCTCGGGCCCTCGCGGAGCTACAGCCCTTACGAGGGTCTCGGAGGGGGGTT	2235		Db	2638	CTGCGCGGCCCTTGGACACCATCAACGTCCACCTCCGGGCTGGGTACATCATCCCCCTG	2697	
Db	1558	AGCAGCTCGGGCCCTCGCGGAGCTACAGCCCTTACGAGGGTCTCGGAGGGGGTT	1617		QY	3316	CAGGGCCCTTGGCTTCAACCAACAGAGTCCCGCCAGAGCCCATGGCCCTTGGCTGTGGCC	3375	
QY	2236	TTTCATCAACAACGAGACCGGCCAGCCGCTGATTGGGAAGGTATGGGCCCGGGTCCACTGCC	2295						


```
Db 1839 TTTCATCACCAACGAGACCGGCGCTGATTGGGAAGTATGGCCCGGCTCCACTGCC 1898
QY 2296 TTCCCCGACTTACCAACCCCAAGCCCTGGCTGGCTGGAGACATGCTGCTGAGTTC 2355
Db 1899 TTCCCCGACTTACCAACCCCAAGCCCTGGCTGGCTGGAGACATGCTGCTGAGTTC 1958
QY 2356 CATGACAGGTGCTTCGACGGCTTGTGGAATTGACATGAACGAGCCCTTCCAACTTCATC 2415
Db 1959 CATGACAGGTGCTTCGACGGCTTGTGGAATTGACATGAACGAGCCCTTCCAACTTCATC 2018
QY 2416 AGAGGCTCTGAGGACGGCTGCCCAACATGAGCTGGAGAAACCCACCTTACGTCCTGG 2475
Db 2019 AGGGGCTCTGAGGACGGCTGCCCAACATGAGCTGGAGAAACCCACCTTACGTCCTGG 2078
QY 2476 GTGGTTGGGGGGACCTCCAGGGGGCCACCATCTGCTCCAGCCACCAAGTTTCTCTCC 2535
Db 2079 GTGGTTGGGGGGACCTCCAGGGGGCCACCATCTGCTCCAGCCACCAAGTTTCTCTCC 2138
QY 2536 ACACACTACAACCTGCACAACTCTACGGCTGACCGAAGCCATCGCTCCACAGGGCG 2595
Db 2139 ACACACTACAACCTGCACAACTCTACGGCTGACCGAAGCCATCGCTCCACAGGGCG 2198
QY 2596 CTGGTGAAGGCTCGGGGACAGGCCCATTTGTGATCTCCGGTCTGACCTTTGCTGGCCAC 2655
Db 2199 CTGGTGAAGGCTCGGGGACAGGCCCATTTGTGATCTCCGGTCTGACCTTTGCTGGCCAC 2258
QY 2656 GCGCGATACGCGGCCACTGAGGGGGAGCTGTGGAGCTCTTGGAGCAGCTCGCCTCC 2715
Db 2259 GCGCGATACGCGGCCACTGAGGGGGAGCTGTGGAGCTCTTGGAGCAGCTCGCCTCC 2318
QY 2716 TCCGTGCCAGAAATCTCGAGTTTAACTCTGCTGGGGGTGCTGTGGTGGGGCCAGCGTC 2775
Db 2319 TCCGTGCCAGAAATCTCGAGTTTAACTCTGCTGGGGGTGCTGTGGTGGGGCCAGCGTC 2378
QY 2776 TGCGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGGTGGTGGACCCAGCTGGGGGCC 2835
Db 2379 TGCGGCTTCTTGGGCAACCTCAGAGGAGCTGTGTGGTGGTGGACCCAGCTGGGGGCC 2438
QY 2836 TTCTACCTCTCATGCGGACCAACACAGCTGCTCAGTCTGCCAGGAGCGGTACAGC 2895
Db 2439 TTCTACCTCTCATGCGGACCAACACAGCTGCTCAGTCTGCCAGGAGCGGTACAGC 2498
QY 2896 TTCCAGGAGCGCGCCAGCAGGCGCATGAGGAAGGCCCTCACCTGGCTGCTACGACTCTTC 2955
Db 2499 TTCCAGGAGCGCGCCAGCAGGCGCATGAGGAAGGCCCTCACCTGGCTGCTACGACTCTTC 2558
QY 2956 CCCACCTCTACACACTGTTTCCACAGGCGCAGCTCGGGGGAGACCGTGGCCCGGCC 3015
Db 2559 CCCACCTCTACACACTGTTTCCACAGGCGCAGCTCGGGGGAGACCGTGGCCCGGCC 2618
QY 3016 CTCTTCTCGAGTTTCCCAAGGACTCTAGCACCTGGACTGTGGACCAACAGCTCTGTGG 3075
Db 2619 CTCTTCTCGAGTTTCCCAAGGACTCTAGCACCTGGACTGTGGACCAACAGCTCTGTGG 2678
QY 3076 GGGGAGCGCCCTGCTCATACCCAGTGTCTCAGGCGGGAGCGCCGAAAGTACTGGCTAC 3135
Db 2679 GGGGAGCGCCCTGCTCATACCCAGTGTCTCAGGCGGGAGCGCCGAAAGTACTGGCTAC 2738
QY 3136 TTCCCTTGGGCAATGTTGACACTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTTC 3195
Db 2739 TTCCCTTGGGCAATGTTGACACTGACAGCGGTGCCAATAGAGCCCTTGGCAGCCTTC 2798
QY 3196 CCACCCCACTGACGCTCCGCTGAGCCAGCCATCCAGCAGGAGGGCAGTGGGTGAGC 3255
Db 2799 CCACCCCACTGACGCTCCGCTGAGCCAGCCATCCAGCAGGAGGGCAGTGGGTGAGC 2858
QY 3256 CTGCGGGCGCCCTTGACACCAATCAACAGTCCACCTCCGGGCTGGGTACATCATCCCCCTG 3315
Db 2859 CTGCGGGCGCCCTTGACACCAATCAACAGTCCACCTCCGGGCTGGGTACATCATCCCCCTG 2918
QY 3316 CAGGGCCCTGCGCTCTCAACACACAGAGTCCCGCCAGACCCCATGCGCCCTGCTGGTGGCC 3375
```

```
Db 2919 CAGGGCCCTGGCCTCACAAACACAGAGTCCCGCCAGCAGCCCATGGCCCTGCTGTGGCC 2978
QY 3376 CTAAACAAAGGGTGGAGAGGCCCGAGGGAGCTGTTCTGGAGCATGGAGAGCCTGGAA 3435
Db 2979 CTGACCAAGGGTGGAGAGGCCCGAGGGAGCTGTTCTGGAGCATGGAGAGCCTGGAA 3038
QY 3436 GTGCTGGAGCGAGGGGCCCTACACAGAGTCAATCTTCTGGCCGAGGAATAACAGATCGTG 3495
Db 3039 GTGCTGGAGCGAGGGGCCCTACACAGAGTCAATCTTCTGGCCGAGGAATAACAGATCGTG 3098
QY 3496 AATGAGCTGTTACGTGTGACCAAGTGGAGAGCTGCGCTGCAGCTCAGNAGGTGACTGTC 3555
Db 3099 AATGAGCTGTTACGTGTGACCAAGTGGAGAGCTGCGCTGCAGCTCAGNAGGTGACTGTC 3158
QY 3556 CTGGGCGTGGCCACAGCGGCCCCAGCAGGTCTCTTCCAAAGGTGCTGCTCTCCAACTTC 3615
Db 3159 CTGGGCGTGGCCACAGCGGCCCCAGCAGGTCTCTTCCAAAGGTGCTGCTCTCCAACTTC 3218
QY 3616 ACCTACAGCCCGACACCAAGGTCTTGACATCTGTGCTGCTGCTGATGAGGAGAGCAG 3675
Db 3219 ACCTACAGCCCGACACCAAGGTCTTGACATCTGTGCTGCTGCTGATGAGGAGAGCAG 3278
QY 3676 TTTCTCGTCAGCTGCTGTT 3694
Db 3279 TTTCTCGTCAGCTGCTGTT 3297

RESULT 5
ADF47520
ID ADF47520 standard; cDNA; 2851 BP.
XX AC ADF47520;
XX DT 26-FEB-2004 (first entry)
XX DE Human GAA encoding cDNA SEQ ID NO:23.
XX KW underglycosylated targeted therapeutic; human; lysosome;
KW lysosomal targeting domain;
KW cation-independent mannose-6-phosphate receptor;
KW underglycosylated therapeutic fusion protein; nephrotropic;
KW enzyme replacement therapy; cell therapy; gene therapy;
KW lysosomal storage disease; metabolic disease; enzyme defect;
KW Pompe disease; Tay-Sachs disease; Sandhoff disease; Fabry disease;
KW Gaucher disease; Krabbe disease; Wolman disease; Hurler syndrome;
KW Hunter syndrome; Sly syndrome; Schindler disease;
KW infantile sialic acid storage disease; Batten disease;
KW infantile neuronal ceroid lipofuscinosis; Ehlers-Danlos syndrome type VI;
KW congenital glycosylation disorder; gene; ss.
XX OS Homo sapiens.
XX PN WO2003102583-A1.
XX XX 11-DEC-2003.
XX PD 29-MAY-2003; 2003WO-US017211.
XX PF 29-MAY-2002; 2002US-0384452P.
XX PR 05-JUN-2002; 2002US-0386019P.
XX PR 06-SEP-2002; 2002US-0408816P.
XX PR 16-OCT-2002; 2002US-00272531.
XX PR 06-FEB-2003; 2003US-0445734P.
XX (SYMB-) SYMBIONICS INC.
XX PA Lebowitz JH, Beverley SM, Sly WS;
XX PI WPI; 2004-035473/03.
XX DR P-PSDB; ADF47521.
XX PT Novel underglycosylated targeted therapeutic comprising therapeutic agent
active in human lysosome, lysosomal targeting domain binds to human
```

cation-independent mannose-6-phosphate receptor, for treating Pompe disease.

Example 13B; SEQ ID NO 23; 137pp; English.

The present invention describes an underglycosylated targeted therapeutic GT which comprised: (a) a therapeutic agent that is therapeutically active in human lysosome; and (b) a lysosomal targeting domain that binds an extracellular domain of human cation-independent mannose-6-phosphate receptor (CM) and (i) does not bind a mutain in which amino acid 1572 of CM is changed from isoleucine to threonine; and (ii) binds the mutain with dissociation constant at least ten times the dissociation constant for binding CM. Also described: (1) an underglycosylated therapeutic fusion protein (FP) comprising a therapeutic domain and a subcellular targeting domain that binds to an extracellular domain of a receptor on an exterior surface of a cell, and upon internalisation of the receptor, permits localisation of the therapeutic domain to a subcellular compartment where the therapeutic domain is therapeutically active; and (2) production of GT. GT has nephrotropic activity, and can be used in enzyme replacement therapy, cell therapy and gene therapy. The FP can be used for treating a lysosomal storage disease patient by administering FP to the patient. GT can be used for treating a patient by identifying a targeting moiety that binds CM in a mannose-6-phosphate independent manner, synthesising GT comprising a therapeutic agent that is therapeutically active in a mammalian lysosome and a targeting moiety that binds CM in a mannose-6-phosphate independent manner and administering GT to the patient, where the targeting moiety is identified by screening a nucleic acid or peptide library. GT is useful for treating metabolic disease, lysosomal storage diseases and associated enzyme defects such as Pompe disease, Tay-Sachs disease, Sandhoff disease, Fabry disease, Gaucher disease, Krabbe disease, Wolman disease, Hurler syndrome, Hunter syndrome, Sly syndrome, Schindler disease, infantile sialic acid storage disease, Batten disease, infantile neuronal ceroid lipofuscinosis, Ehlers-Danlos syndrome type VI and congenital disorders of glycosylation. The present sequence is used in the exemplification of the present invention.

Sequence 2851 BP; 510 A; 1001 C; 840 G; 500 T; 0 U; 0 Other;

Very Match	71.2%;	Score 2635.4;	DB 12;	Length 2851;
st Local Similarity	99.6%;	Pred. No. 0;		
atches 2642;	Conservative	0;	Mismatches 11;	Indels 0; Gaps 0;
1036	GAACCGGTGCACACCCCGCGCGTCCACAGACAGTGCACACAGTGCAGCTCCCCCCC	1095		
199	GATGCCAGGCACACCCCGCGCGTCCAGAGAGTGCACACAGTGGAGCTCCCCCCC	258		
1096	AACAGCCGCTTCGATTGGCGCCCTGCAAGCGCCATCACCCAGGAAACAGTGCAGGCGCCGC	1155		
259	AACAGCCGCTTCGATTGGCGCCCTGCAAGGCCATCACCCAGGAAACAGTGCAGGCGCCGC	318		
1156	GGCTGCTGTATATCCCTGCAAGCAGGGGCTGCAGGGAGCCACAGATGGGCGACCTGTG	1215		
319	GGCTGCTGTATATCCCTGCAAGCAGGGGCTGCAGGGAGCCACAGATGGGCGACCTGTG	378		
1216	TGCTTTCTTCCCAACCCAGCTACCCAGCTACAAGCTGGAGAACCCTGAGCTCTCTGAAATG	1275		
379	TGCTTTCTTCCCAACCCAGCTACCCAGCTACAAGCTGGAGAACCCTGAGCTCTCTGAAATG	438		
1276	GGCTACAGGGCCACCTGACCCGTACCAACCCCACTTTCTTCCCAAGGACATCTCGACC	1335		
439	GGCTACAGGGCCACCTGACCCGTACCAACCCCACTTTCTTCCCAAGGACATCTCGACC	498		
1336	CTGCGGCTGGAGCTGATGATGGAGACTGAGAACCGCTCCACTTCAAGATCAAGATCCA	1395		
499	CTGCGGCTGGAGCTGATGATGGAGACTGAGAACCGCTCCACTTCAAGATCAAGATCCA	558		
1396	GCTAACAGCGCTACGAGGTGCCCTTGGAGACCCCGCGTGTCCACAGCCGGGCAACCGTCC	1455		
559	GCTAACAGCGCTACGAGGTGCCCTTGGAGACCCCGCGTGTCCACAGCCGGGCAACCGTCC	618		
1456	CCACTCTACGCGTGGAGTTCTCGAGAGGCCCTTCCGGGTGATCGTGCAACCGCAGCTG	1515		

619	DB	CCACTCTACAGCGTGAGTTCCTCTGAGGAGCCCTTCGGGGTGATCGTGACACCGCAGCTG	678
1516	QY	GACGGCGCGTGTCTGTGAACACGACGGTGGCCCTCTGTTCCTTTGCGGACCAAGTTCCCTT	1575
679	DB	GACGGCGCGTGTCTGTGAACACGACGGTGGCCCTCTGTTCCTTTGCGGACCAAGTTCCCTT	738
1576	QY	CAGCTGTCCACCTGCTGCCCTCGAGTATATACAGGCGCTCGCGGACCACTCAGTCC	1635
739	DB	CAGCTGTCCACCTGCTGCCCTCGAGTATATACAGGCGCTCGCGGACCACTCAGTCC	798
1636	QY	CTGATGCTCAGCACCACTGGACACGATCACCCCTGTGGAAACCGGGACCTTGGCGCCACG	1695
799	DB	CTGATGCTCAGCACCACTGGACACGATCACCCCTGTGGAAACCGGGACCTTGGCGCCACG	858
1696	QY	CCCGTTCGAAACCTCTACCGGTCACCCCTTTCTACTGCGCTGGAGGACGCGGGTCG	1755
859	DB	CCCGTTCGAAACCTCTACCGGTCACCCCTTTCTACTGCGCTGGAGGACGCGGGTCG	918
1756	QY	GCACAGGGGTGTTCTGTCTAAACAGCAATGCCATGGATGTGTCTCTGACCGCGGCCCT	1815
919	DB	GCACAGGGGTGTTCTGTCTAAACAGCAATGCCATGGATGTGTCTCTGACCGCGGCCCT	978
1816	QY	GCCCTTAGCTGGAGGTGCGACAGGTGGATCCTCGATGTCTACATCTTCTCTGGGCCACAG	1875
979	DB	GCCCTTAGCTGGAGGTGCGACAGGTGGATCCTCGATGTCTACATCTTCTCTGGGCCACAG	1038
1876	QY	CCCAAGAGCGTGTGACAGTAGTACCTGGAGTTGTGGGATACCCTGTTCTATGCCGCCATAC	1935
1039	DB	CCCAAGAGCGTGTGACAGTAGTACCTGGAGTTGTGGGATACCCTGTTCTATGCCGCCATAC	1098
1936	QY	TGGGGCTTGGCTTCCACTGTGCGCTGGGGTACTCTCTCACCGCTATCACCCGCCAG	1995
1099	DB	TGGGGCTTGGCTTCCACTGTGCGCTGGGGTACTCTCTCACCGCTATCACCCGCCAG	1158
1996	QY	GTGGTGGAGAAATGACACAGGGCCCACTTCCCTCGACGTCCAAATGGAAACGACTCGAC	2055
1159	DB	GTGGTGGAGAAATGACACAGGGCCCACTTCCCTCGACGTCCAAATGGAAACGACTCGAC	1218
2056	QY	TACATGACTCCCGGAGGACCTTACGTTCAAACAGGATGGCTTCCGGGACTTCCCGGCC	2115
1219	DB	TACATGACTCCCGGAGGACTTACGTTCAAACAGGATGGCTTCCGGGACTTCCCGGCC	1278
2116	QY	ATGGTCAGAGCTGCACAGGGCGCGCGCTACATGATGATCGTGGATCTGCGCATC	2175
1279	DB	ATGGTCAGAGCTGCACAGGGCGCGCGCTACATGATGATCGTGGATCTGCGCATC	1338
2176	QY	AGCAGCTCGGCGCTCGCGGAGCTACAGCCCTTACGACGAGGCTTGGCGGAGGGGTT	2235
1339	DB	AGCAGCTCGGCGCTCGCGGAGCTACAGCCCTTACGACGAGGCTTGGCGGAGGGGTT	1398
2236	QY	TTCAATCAACAGACCCGCCACGCGCTGATTTGGGAAAGGTATGGCCCGGGTCCACTGCC	2295
1399	DB	TTCAATCAACAGACCCGCCACGCGCTGATTTGGGAAAGGTATGGCCCGGGTCCACTGCC	1458
2296	QY	TTCCCCGACTTCAACAAACCCCAAGCCCTGCTGTGTGGAGGACATGTGTGGCTGAGTTC	2355
1459	DB	TTCCCCGACTTCAACAAACCCCAAGCCCTGCTGTGTGGAGGACATGTGTGGCTGAGTTC	1518
2356	QY	CATGACGAGGTGCCCTTCGACGGCTTGTGGATGACATGAACAGGCTTCCAACTTCATC	2415
1519	DB	CATGACGAGGTGCCCTTCGACGGCATGTGGATGACATGAACAGGCTTCCAACTTCATC	1578
2416	QY	AGAGGCTCTGAGGACGGCTGCCCAACATATGACTGGAGAACCCACTACGTGCTCTGGG	2475
1579	DB	AGGGGCTCTGAGGACGGCTGCCCAACATATGACTGGAGAACCCACTACGTGCTCTGGG	1638
2476	QY	GTGGTTGGGGGACCTTCCAGGGCGCAACATCTGTGCTCTCCAGCAACCAAGTTTCTCTCC	2535
1639	DB	GTGGTTGGGGGACCTTCCAGGGCGCAACATCTGTGCTCTCCAGCAACCAAGTTTCTCTCC	1698
2536	QY	ACACACTAACCTGCACAACTTACGGCTGACCGAGCCATCGCTCTCCACAGGGCG	2595
1699	DB	ACACACTAACCTGCACAACTTACGGCTGACCGAGCCATCGCTCTCCACAGGGCG	1758

```
QY 2596 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTTGTGCGCAC 2655
Db 1759 CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTTGTGCGCAC 1818
QY 2656 GGCAGATACGCGGCCACATCGAGCGGGGAGCTGTGGAGTCTCTGGAGCAGCTCGCTCC 2715
Db 1819 GGCAGATACGCGGCCACATCGAGCGGGGAGCTGTGGAGTCTCTGGAGCAGCTCGCTCC 1878
QY 2716 TCCGTGCCAGAAATCTCGAGTTTAACTCTGGGGGTCCCTCTGTCGGGGCCGACGTC 2775
Db 1879 TCCGTGCCAGAAATCTCGAGTTTAACTCTGGGGGTCCCTCTGTCGGGGCCGACGTC 1938
QY 2776 TCGGCTTCTCGGCAACACCTCAGAGGAGCTGTGTGCGCTGGACCCAGCTGGGGGCC 2835
Db 1939 TCGGCTTCTCGGCAACACCTCAGAGGAGCTGTGTGCGCTGGACCCAGCTGGGGGCC 1998
QY 2836 TTCTACCTCTTATGCGGAACACACAGCTGTCTCAGTCTGCTCCAGAGCGGTACAGC 2895
Db 1999 TTCTACCTCTTATGCGGAACACACAGCTGTCTCAGTCTGCTCCAGAGCGGTACAGC 2058
QY 2896 TTCAGGAGCGCGCCAGCAGGCCATGAGAGGGCCCTCACCTCGCTACGACTCTCTC 2955
Db 2059 TTCAGGAGCGCGCCAGCAGGCCATGAGAGGGCCCTCACCTCGCTACGACTCTCTC 2118
QY 2956 CCCCACCTCTACACACTGTTCCACACGAGCCACGTCGCGGGGAGACCGTGGCCCGGCC 3015
Db 2119 CCCCACCTCTACACACTGTTCCACACGAGCCACGTCGCGGGGAGACCGTGGCCCGGCC 2178
QY 3016 CTCCTCTCGAGTTCGCCAAGACTCTAGCACTGTGACCTGTGGACACACAGCTCTGTGG 3075
Db 2179 CTCCTCTCGAGTTCGCCAAGACTCTAGCACTGTGACCTGTGGACACACAGCTCTGTGG 2238
QY 3076 GGGAGGCGCTCTCATATCCCGAGTCCTCAGGCGGGAAGCCGGAAGTACTGCTAC 3135
Db 2239 GGGAGGCGCTCTCATATCCCGAGTCCTCAGGCGGGAAGCCGGAAGTACTGCTAC 2298
QY 3136 TTCCCTCTGGGCACATGTCAGCACTGCAGACGGTGCCTAATAGAGCCCTTGGCAGCTTC 3195
Db 2299 TTCCCTCTGGGCACATGTCAGCACTGCAGACGGTGCCTAATAGAGCCCTTGGCAGCTTC 2358
QY 3196 CACCCCGCTCTGAGCTCCCGTGTAGCCAGCCATCCACAGGAGGGGAGTGGGTGAGC 3255
Db 2359 CACCCCGCTCTGAGCTCCCGTGTAGCCAGCCATCCACAGGAGGGGAGTGGGTGAGC 2418
QY 3256 CTGCGGCGCCCTGTGACACCATCAACGTCACCTCCGCGCTGGGTACATCATCCCCCTG 3315
Db 2419 CTGCGGCGCCCTGTGACACCATCAACGTCACCTCCGCGCTGGGTACATCATCCCCCTG 2478
QY 3316 CAGGGCCCTGGCTCACAACCCAGAGTCCCGCCAGCAGCCCATGCGCTGTGGCC 3375
Db 2479 CAGGGCCCTGGCTCACAACCCAGAGTCCCGCCAGCAGCCCATGCGCTGTGGCC 2538
QY 3376 CTAAACAAAGGTGAGAGGCGCCAGGGGAGCTGTCTGGGACGATGGAGAGAGCTGGAA 3435
Db 2539 CTAAACAAAGGTGAGAGGCGCCAGGGGAGCTGTCTGGGACGATGGAGAGAGCTGGAA 2598
QY 3436 GTGCTGAGCGAGGGGCTTACACACAGGTCATCTCTCGCCAGGAATTAACAGATCTGTG 3495
Db 2599 GTGCTGAGCGAGGGGCTTACACACAGGTCATCTCTCGCCAGGAATTAACAGATCTGTG 2658
QY 3496 AATGAGCTGTGTAGCTGACAGTGAAGGAGCTGGCTCAGCTGCAGAGGTGACTGTTC 3555
Db 2659 AATGAGCTGTGTAGCTGACAGTGAAGGAGCTGGCTCAGCTGCAGAGGTGACTGTTC 2718
QY 3556 CTGGGCGTGGCCACGCGGCCCCAGCAGGTCCTCTCCAAAGGTGTCTCTCTCCAACTTC 3615
Db 2719 CTGGGCGTGGCCACGCGGCCCCAGCAGGTCCTCTCCAAAGGTGTCTCTCTCCAACTTC 2778
QY 3616 ACCTACAGCCCGACACCAAGGTCCTGGACATCTGTGCTCGCTGTGGAGAGCAG 3675
Db 2779 ACCTACAGCCCGACACCAAGGTCCTGGACATCTGTGCTCGCTGTGGAGAGCAG 2838
```

```
QY 3676 TTTCTCGTCAGCT 3688
Db 2839 TTTCTCGTCAGCT 2851

RESULT 6
ACN42967
ID ACN42967 standard; cDNA; 3376 BP.
XX
AC ACN42967;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide SEQ ID NO:1842.
KW ss; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
XX dithp.
XX Homo sapiens.
OS
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Riddy TP;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson JB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton BS;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
WPI: 2004-329368/30.
DR P-PSDB; ABM84315.
XX
New diagnostic and therapeutic polynucleotides and polypeptides, useful
in diagnosing a condition, disease or disorder associated with human
molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
in gene mapping.
XX
Claim 1; Page; 190pp; English.
XX
The invention relates to novel diagnostic and therapeutic polynucleotides
selected from one of the 2722 sequences defined in the specification. A
polynucleotide of the invention may have a use in gene therapy. The human
diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
used to diagnose a particular condition, disease or disorder associated
with human molecules, e.g. cell proliferative disorders,
autoimmune/inflammatory disorder, developmental disorder, endocrine
disorder, neurological disorders, gastrointestinal disorders, or
infections caused by virus, bacteria, fungi or parasite. The dithp
molecules may also be used in genetic mapping, in identifying individuals
from minute biological samples, in detecting single nucleotide
polymorphisms, as molecular weight markers, and for somatic or germline
gene therapy. The present sequence represents a dithp polynucleotide of
the invention. Note: The sequence data for this patent is not represented
in the printed specification, but was obtained in electronic format
directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
Sequence 3376 BP; 590 A; 1153 C; 1020 G; 613 T; 0 U; 0 Other;
Query Match 59.1%; Score 2186.2; DB 13; Length 3376;
Best Local Similarity 91.2%; Pred. No. 0;
Matches 2425; Conservative 0; Mismatches 13; Indels 221; Gaps 1;
```

QY	1036	GAACCGGTGCACACCCCGGCGTCCAGAGAGAGTGCCCAACACAGTGCAGAGTCCCGCC	1095
DB	454		
QY	1096	GATGCCCAGGCAACACCCCGGCGTCCAGAGAGAGTGCCCAACAGTGCAGAGTCCCGCC	513
DB	514		
QY	1156	AACAGCGGCTTCGATTGCGCCCTGACAGAGGCCATCACACAGAAACAGTGCAGAGGCCGC	1155
DB	574		
QY	1216	AACAGCGGCTTCGATTGCGCCCTGACAGAGGCCATCACACAGAAACAGTGCAGAGGCCGC	573
DB	634		
QY	1276	GGCTGCTGTACATCCCTGCAAGAGGGGCTGACAGAGGCCAGATGGGCGAGCCCTGG	1215
DB	694		
QY	1336	GGCTGCTGTACATCCCTGCAAGAGGGGCTGACAGAGGCCAGATGGGCGAGCCCTGG	633
DB	754		
QY	1396	TGCTTCTTCCACCGAGCTACCCAGCTACAGAGGCCAGATGGGCGAGCCCTGG	1275
DB	814		
QY	1456	TGCTTCTTCCACCGAGCTACCCAGCTACAGAGGCCAGATGGGCGAGCCCTGG	693
DB	934		
QY	1516	GGCTACACGGCCACCTGACCGCTTACACCCCGAGCTTCTTCCCAAGGACATCTGACC	1335
DB	994		
QY	1576	GGCTACACGGCCACCTGACCGCTTACACCCCGAGCTTCTTCCCAAGGACATCTGACC	753
DB	1054		
QY	1636	CTCGGCTGGAAGTGTGATGGAGACTGAGAACCGGCTTCCACTTCAAGATCCA	1395
DB	1114		
QY	1696	CTCGGCTGGAAGTGTGATGGAGACTGAGAACCGGCTTCCACTTCAAGATCCA	813
DB	1174		
QY	1756	CTCGGCTGGAAGTGTGATGGAGACTGAGAACCGGCTTCCACTTCAAGATCCA	813
DB	1234		
QY	1816	GCTACACGGCGCTACAGAGTGCCCTTGGAGACCCCGGCTGTCCACAGCCGGGACCGCTCC	1455
DB	1294		
QY	1876	GCTACACGGCGCTACAGAGTGCCCTTGGAGACCCCGGCTGTCCACAGCCGGGACCGCTCC	873
DB	1354		
QY	1936	CCACTCTACAGCGTGAGTCTCCGAGGAGCCCTTCGGGGTGATCGTGACCGGAGCTG	933
DB	1414		
QY	1996	CCACTCTACAGCGTGAGTCTCCGAGGAGCCCTTCGGGGTGATCGTGACCGGAGCTG	933
DB	2056		
QY	2076	GACGGCGGGTGCTGTGAACACGAGCGTGCGGCCCTTCTTTCGGGACACAGTTCCCTT	1575
DB	2116		
QY	2176	GACGGCGGGTGCTGTGAACACGAGCGTGCGGCCCTTCTTTCGGGACACAGTTCCCTT	993
DB	2236		
QY	2296	CHAGCTGTCAAGTGTGCTGTGAACACGAGCGTGCGGCCCTTCTTTCGGGACACAGTTCCCTT	1635
DB	2356		
QY	2416	CHAGCTGTCAAGTGTGCTGTGAACACGAGCGTGCGGCCCTTCTTTCGGGACACAGTTCCCTT	1053
DB	2476		
QY	2536	CTGATGCTCAGACACGAGCTGGACAGGATCACCTGTGGAACCGGAGACCTTGGGCCACG	1695
DB	2596		
QY	2656	CTGATGCTCAGACACGAGCTGGACAGGATCACCTGTGGAACCGGAGACCTTGGGCCACG	1113
DB	2716		
QY	2776	CCCGGTGCAACCTCTACCGGCTCTACACCTTCTACCTGGCGTGGAGACCGCGGGTGG	1755
DB	2836		
QY	2896	CCCGGTGCAACCTCTACCGGCTCTACACCTTCTACCTGGCGTGGAGACCGCGGGTGG	1173
DB	2956		
QY	3016	GCACACGGGGTGTCTGTCTAAACAGCAATGCGATGGATGTGTCTGACGCGGAGCCCT	1815
DB	3076		
QY	3136	GCACACGGGGTGTCTGTCTAAACAGCAATGCGATGGATGTGTCTGACGCGGAGCCCT	1233
DB	3196		
QY	3256	GCCTTAGCTGGAGGTGCAGGTGGAGTCTGGATGTCTACATCTTCTGGGCCACAG	1875
DB	3316		
QY	3376	GCCTTAGCTGGAGGTGCAGGTGGAGTCTGGATGTCTACATCTTCTGGGCCACAG	1293
DB	3436		
QY	3496	CCCAAGAGCGTGTGCAGAGTACCTTGGAGTGTGGGATACCCGTTCAATGCGGCCATAC	1935
DB	3556		
QY	3616	CCCAAGAGCGTGTGCAGAGTACCTTGGAGTGTGGGATACCCGTTCAATGCGGCCATAC	1353
DB	3676		
QY	3736	TGGGGCTTGGGCTTCACTGTGCGGCTGAGGCTACTCTCTCAACCGCTATACCGGCCAG	1995
DB	3796		
QY	3856	TGGGGCTTGGGCTTCACTGTGCGGCTGAGGCTACTCTCTCAACCGCTATACCGGCCAG	1413
DB	3916		
QY	3976	GTGGTGGAGAACATGACAGGGGCCACTTCCCGCTGGACGTCCCAATGGAAACGACTGGAC	2055
DB	4036		
QY	4096	GTGGTGGAGAACATGACAGGGGCCACTTCCCGCTGGACGTCCCAATGGAAACGACTGGAC	1473
DB	4156		
QY	4216	TACATGGACTCCCGAGGAGACTTACGTTCAACAGGATGGCTTCCGGGACTTCCCGGCC	2115
DB	4276		
QY	4336	TACATGGACTCCCGAGGAGACTTACGTTCAACAGGATGGCTTCCGGGACTTCCCGGCC	1533
DB	4396		
QY	4456	ATGGTGCAGGAGCTGCACACGAGCGCGCGCTACATGATGATGATGATGATGATGATGATG	2175

DB	1534	ATGGTGCAGGAGCTGCACACGAGCGCGCGCTACATGATGATGATGATGATGATGATGATG	1593
QY	2176	AGCAGCTCGGGCCCTGCGCGGAGCTACAGCCCTACGACAGAGGTCTGCGAGGGGGGTT	2235
DB	1594	AGCAGCTCGGGCCCTGCGCGGAGCTACAGCCCTACGACAGAGGTCTGCGAGGGGGGTT	1653
QY	2236	TTTCATCAACAAAGAGACCGGCGCTGATTTGGAAAGTATGGCCCGGGTCCATGCG	2295
DB	1654	TTTCATCAACAAAGAGACCGGCGCTGATTTGGAAAGTATGGCCCGGGTCCATGCG	1713
QY	2296	TTTCATCAACAAAGAGACCGGCGCTGATTTGGAAAGTATGGCCCGGGTCCATGCG	2355
DB	1714	TTTCATCAACAAAGAGACCGGCGCTGATTTGGAAAGTATGGCCCGGGTCCATGCG	1773
QY	2356	CATGACCAAGTCTTCCAGCGGCTTGGATTTGACATGAACGAGGCTTCCAACTTCATC	2415
DB	1774	CATGACCAAGTCTTCCAGCGGCTTGGATTTGACATGAACGAGGCTTCCAACTTCATC	1833
QY	2416	AGAGGCTCTGAGGACGGCTGCCCCAAACATGATGCTGGAGAACCCACCTTACGTGCGTGG	2475
DB	1834	AGAGGCTCTGAGGACGGCTGCCCCAAACATGATGCTGGAGAACCCACCTTACGTGCGTGG	1893
QY	2476	GTGGTGGGGGACCGCTCCAGCGGCGCACCATCTGTGCCTCCAGCCACAGTTTCTCTCC	2535
DB	1894	GTGGTGGGGGACCGCTCCAGCGGCGCACCATCTGTGCCTCCAGCCACAGTTTCTCTCC	1953
QY	2536	ACACACTACAACTCTGACAACTCTACGCGCTGACGAGGACCATCGCTTCCACAGGGCG	2595
DB	1954	ACACACTACAACTCTGACAACTCTACGCGCTGACGAGGACCATCGCTTCCACAGGGCG	2013
QY	2596	CTGGTAAAGGCTCGGGGACACCGCCATTTGTGATCTCCCGCTCGACCTTTGTGCGCCAC	2655
DB	2014	CTGGTAAAGGCTCGGGGACACCGCCATTTGTGATCTCCCGCTCGACCTTTGTGCGCCAC	2073
QY	2656	GGCGATACCGCGGCACTGGAGCGGGGACGTTGGAGCTCTCTGGAGCAGCTCGCGCTCC	2715
DB	2074	GGCGATACCGCGGCACTGGAGCGGGGACGTTGGAGCTCTCTGGAGCAGCTCGCGCTCC	2133
QY	2716	TCGGTGCAGAAATCTTCAGGTTTAACTGCTGGGGGTGCTCTGCTCGGGGCGGACGTC	2775
DB	2134	TCGGTGCAGAAATCTTCAGGTTTAACTGCTGGGGGTGCTCTGCTCGGGGCGGACGTC	2193
QY	2776	TCGGGCTTCTGGGCAACCTCAGAGGAGTGTGTGCTGCTGGACCCAGCTGGGGGCG	2835
DB	2194	TCGGGCTTCTGGGCAACCTCAGAGGAGTGTGTGCTGCTGGACCCAGCTGGGGGCG	2253
QY	2836	TTCTACCCCTTCATGCGGNAACCAACAGCTGCTCAGTCTGCTCCCGAGGCGGTACAGC	2895
DB	2254	TTCTACCCCTTCATGCGGNAACCAACAGCTGCTCAGTCTGCTCCCGAGGCGGTACAGC	2313
QY	2896	TTTCAGGAGCGCGCCAGCAGGACCATGAGAAAGGCTTCACTGCTGCTACGCACTTCTTC	2955
DB	2314	TTTCAGGAGCGCGCCAGCAGGACCATGAGAAAGGCTTCACTGCTGCTACGCACTTCTTC	2373
QY	2956	CCCCACCTTACACACTGTTCACACAGGCCACGTCGCGGGGAGACCGTGGCGCGGCC	3015
DB	2374	CCCCACCTTACACACTGTTCACACAGGCCACGTCGCGGGGAGACCGTGGCGCGGCC	2433
QY	3016	CTCTTCTGGAGTTCGCCAAGGACTTAGCACTGTGGAATGTTGGGCTCTGCTGCTG	3075
DB	2434	CTCTTCTGGAGTTCGCCAAGGACTTAGCACTGTGGAATGTTGGGCTCTGCTGCTG	2470
QY	3076	GGGGAGGCGCTCTCATCATCCCCAGTGTCTCCAGCGCGGAGGCGGAGTGTGCTAC	3135
DB	2471	-----	2470
QY	3136	TTCCCTTTGGGCACTGTGTACGACCTGCAGACGTTGCCAATAGAGGCCCTTGGCAGCTTC	3195
DB	2471	-----	2470
QY	3196	CCACCCCACTGACAGTCTCCCGTGGCGGAGCCATCCACAGCGAGGGGAGTGGGTGACG	3255

PT	New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.
PS	Claim 1; Page; 190pp; English.
XX	The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2722 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, autoimmune/inflammatory disorders, developmental disorder, endocrine disorder, neurological disorders, gastrointestinal disorders, or infections caused by virus, bacteria, fungi or parasite. The dithp molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline gene therapy. The present sequence represents a dithp polynucleotide of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm
SQ	Sequence 3455 BP; 597 A; 1189 C; 1042 G; 627 T; 0 U; 0 Other;
QY	Query Match
DB	Best Local Similarity 56.7%; Score 2097.2; DB 13; Length 3455; Matches 2425; Conservative 0; Mismatches 13; Indels 300; Gaps 2;
QY	1036 GAAACCGGTGACACCCCGCGCTCCAGAGCAGTGCCTCACACAGTGGACGTCCTCCCCC
DB	454 GATGCCAGGACACCCCGCGCTCCAGAGCAGTGCCTCACACAGTGGACGTCCTCCCCC
QY	1096 AACAGCGCTTCGATTGGCGCCCTGACAGGCGCATCCAGGACAGTGGCGAGCCCGC
DB	514 AACAGCGCTTCGATTGGCGCCCTGACAGGCGCATCCAGGACAGTGGCGAGCCCGC
QY	1156 GGCTGCTGTACATCCCTGCAAGCAGGGGCTGCAAGGAGCCAGATGGGGCAGCCCTGG
DB	574 GGCTGCTGTACATCCCTGCAAGCAGGGGCTGCAAGGAGCCAGATGGGGCAGCCCTGG
QY	1216 TGCTTTTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTGAAATG
DB	634 TGCTTTTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTGAAATG
QY	1276 GGCTACAGGCGCCCTGACCCGCTACCCCGCTGCAAGGAGCCAGATGGGGCAGCCCTGG
DB	694 GGCTACAGGCGCCCTGACCCGCTACCCCGCTGCAAGGAGCCAGATGGGGCAGCCCTGG
QY	1336 CTGGGCTGGAGCTGATGAGAGCTGAGAACCGCTCTCACTTCACGATCAAGATCCA
DB	754 CTGGGCTGGAGCTGATGAGAGCTGAGAACCGCTCTCACTTCACGATCAAGATCCA
QY	1396 GCTAACAGGCGCTACAGAGGTGCCCTTTGGAGACCCCGCTGTCACAGCGGCGACCGTCC
DB	814 GCTAACAGGCGCTACAGAGGTGCCCTTTGGAGACCCCGCTGTCACAGCGGCGACCGTCC
QY	1456 CCATCTACAGCGTGGAGTTCCTCGAGGAGCCCTTGGGGTGTATCTGGCACCGGAGCTG
DB	874 CCATCTACAGCGTGGAGTTCCTCGAGGAGCCCTTGGGGTGTATCTGGCACCGGAGCTG
QY	1516 GACGGCGCGTCTGCTGTAACACAGCGGTGGCGCCCTGTTCTTTGGGAGCAGTTCCTT
DB	934 GACGGCGCGTCTGCTGTAACACAGCGGTGGCGCCCTGTTCTTTGGGAGCAGTTCCTT
QY	1576 CAGCTGTCTCACCTCGCTGCCCTTCGACGTATATACAGGCGCTCGCGGAGCACCTCAGTCCC
DB	994 CAGCTGTCTCACCTCGCTGCCCTTCGACGTATATACAGGCGCTCGCGGAGCACCTCAGTCCC
QY	1636 CTGATGCTCAGCAGCTGGACAGGATCACTCTGTGGAAACCGGAGCTTCGGCCCGAG
DB	1054 CTGATGCTCAGCAGCTGGACAGGATCACTCTGTGGAAACCGGAGCTTCGGCCCGAG

1696 CCGGTGCGAACCCTCTACGGGTCTCAACCTTTCTAAGCTGCGTGGAGGACGCGGGTCG 1755
1114 CCGGTGCGAACCCTCTACGGGTCTCAACCTTTCTAAGCTGCGTGGAGGACGCGGGTCG 1173
1756 GCACGGGGTGTCTCTTAACAGCAATGCCATGGATGTGCTCTGAGCCGAGCCCT 1815
1174 GCACGGGGTGTCTCTTAACAGCAATGCCATGGATGTGCTCTGAGCCGAGCCCT 1233
1816 GCCCTTAGCTGAGGTGCGACAGTGGATCCTGGATGTCTACATCTTCTGGGCCAGAG 1875
1234 GCCCTTAGCTGAGGTGCGACAGTGGATCCTGGATGTCTACATCTTCTGGGCCAGAG 1293
1876 CCACAGAGCGTGTGACGAGTACTTGGAGTGTG----- 1911
1294 CCACAGAGCGTGTGACGAGTACTTGGAGTGTG----- 1353
1912 ----- 1916
1354 GCGCCCGCCCAAGGTCCCTCTCCCTCCCTCATGAAGTCGGCGTTGGCTGACAGGATA 1413
1917 CCGGTTTCATGCCGCCATCTAGGGGCTTGGGCTTCCACCTGTGCGCTGGGGCTACTCCTC 1976
1414 CCGGTTTCATGCCGCCATCTAGGGGCTTGGGCTTCCACCTGTGCGCTGGGGCTACTCCTC 1473
1977 CACCGCTATCACCGCCAGGTGGTGGAGAACATGACAGGGGCCACTTCCCTCGGACGT 2036
1474 CACCGCTATCACCGCCAGGTGGTGGAGAACATGACAGGGGCCACTTCCCTCGGACGT 1533
2037 CCNATGGAACGACTGACATGACATGACATCCCGAGGGAGTTCACGTTCAAGAGATGG 2096
1534 CCAGTGGAAACGACTGACATGACATGACATCCCGAGGGAGTTCACGTTCAAGAGATGG 1593
2097 CTTCCGGGACTTCCCGCCATGTGACAGAGTGCACAGGGGCGCGGCTTACATGAT 2156
1594 CTTCCGGGACTTCCCGCCATGTGACAGAGTGCACAGGGGCGCGGCTTACATGAT 1653
2157 GATCGTGGATCTCGCCATCAGCAGCTCGGGCCCTGCGGGAGCTACAGGCCCTACGACGA 2216
1654 GATCGTGGATCTCGCCATCAGCAGCTCGGGCCCTGCGGGAGCTACAGGCCCTACGACGA 1713
2217 GGGTCTCGGGAGGGGGTTCATACCAACAGAGACGGGCCAGCCCTGATGGGAAGGT 2276
1714 GGGTCTCGGGAGGGGGTTCATACCAACAGAGACGGGCCAGCCCTGATGGGAAGGT 1773
2277 ATGGCCCGGGTCCACTGCGCTTCCCGACTTCCACAAACAGAGCCCTGCGGTGGGA 2336
1774 ATGGCCCGGGTCCACTGCGCTTCCCGACTTCCACAAACAGAGCCCTGCGGTGGGA 1833
2337 GGAATGCTGGTGGTGGTTCATGACAGGTGCGCTTCCGACGGCTTGGATGACATGAA 2396
1834 GGAATGCTGGTGGTGGTTCATGACAGGTGCGCTTCCGACGGCTTGGATGACATGAA 1893
2397 CGAGCCTTCCAACTTCATCAGAGGCTCTGAGGACGGCTGCGCCCAACATGAGCTGAGAA 2456
1894 CGAGCCTTCCAACTTCATCAGAGGCTCTGAGGACGGCTGCGCCCAACATGAGCTGAGAA 1953
2457 CCCACCTAGTGCCTGGGGTGGTGGGGGACCTCCAGGGGCGCCACCATCTGCGCTC 2516
1954 CCCACCTAGTGCCTGGGGTGGTGGGGGACCTCCAGGGGCGCCACCATCTGCGCTC 2013
2517 CAGCCACCAAGTTTCTCTCCACACACTACAACTTCGCAACCTCTAGGGCTGACCGAAGC 2576
2014 CAGCCACCAAGTTTCTCTCCACACACTACAACTTCGCAACCTCTAGGGCTGACCGAAGC 2073
2577 CATCGCTCCCAACAGGGCGTGTGAAGGCTCGGGGACAGCGCCATTTGTGATCTCCCG 2636
2074 CATCGCTCCCAACAGGGCGTGTGAAGGCTCGGGGACAGCGCCATTTGTGATCTCCCG 2133
2637 CTCGACCTTGTGGGCGACGGCCGATACGCGCGCACTGGACGGGGAGCTGTGGAGCTC 2696
2134 CTCGACCTTGTGGGCGACGGCCGATACGCGCGCACTGGACGGGGAGCTGTGGAGCTC 2193

RESULT 8
ACN42968
ID ACN42968 standard; cDNA; 3105 BP.
XX

2697 CTGGAGAGCAGCTCGCTCCTCCGTGCGAAGAAATCTCGAGTTTAACTGCTGGGGGTGCC 2756
2194 CTGGAGAGCAGCTCGCTCCTCCGTGCGAAGAAATCTCGAGTTTAACTGCTGGGGGTGCC 2253
2757 TCTGGTTCGGGGCGGAGCTGCTGGGCTTCTGGGCAACACCTCAGAGAGCTGTGTGTGCG 2816
2254 TCTGGTTCGGGGCGGAGCTGCTGGGCTTCTGGGCAACACCTCAGAGAGCTGTGTGTGCG 2313
2817 CTGGACCCAGCTGGGGGCTTCTTACCCCTTATGCGGAAACCAACAGAGCTGCTCAGTCT 2876
2314 CTGGACCCAGCTGGGGGCTTCTTACCCCTTATGCGGAAACCAACAGAGCTGCTCAGTCT 2373
2877 GCGCCAGAGCGCTGACAGCTTCAAGAGCGCGCCAGAGGCGCATAGAGAGCCCTCAG 2936
2374 GCGCCAGAGCGCTGACAGCTTCAAGAGCGCGCCAGAGGCGCATAGAGAGCCCTCAG 2433
2937 CTTGGCGTACGCACTCTCCCGCACCTTACACACTGTTTCCACAGGCGCCAGCTGCGGG 2996
2434 CTTGGCGTACGCACTCTCTCCCGCACCTTACACACTGTTTCCACAGGCGCCAGCTGCGGG 2493
2997 GGAGACCGTGGGCCCGGCCCTTCTTCTGGAGTTCCTCCAAAGGACTCTAGCACCTGTG 3056
2494 GGAGACCGTGGGCCCGGCCCTTCTTCTGGAGTTCCTCCAAAGGACTCTAGCACCTGGA --- 2549
3057 GGACCAACAGCTCTGTGGGGGGAGGCCCTGCTCATCACCCAGTCTCCAGGCCCGGAA 3116
2550 ----- 2549
3117 GCGCGAAGTACTGGCTACTTCCCTTGGGCACATGGTACGACCTGACAGCGGTGCCAAT 3176
2550 ----- 2549
3177 AGAGGCCCTTGGCAGCCTCCACCCACCTGCGAGCTCCCGGTGAGCCAGCATCCACAG 3236
2550 ----- 2549
3237 CGAGGGGAGTGGGTGACCGCTCGCGGCCCTTGACACCATCAAGTCCACCTCGGGC 3296
2550 ----- ACCATCAAGTCCACCTCGGGC 2572
3297 TGGGTACATCATCCCCCTGCGAGGCCCTGGGCTTCAACACAGAGTCCCGCCAGCAGCC 3356
2573 TGGGTACATCATCCCCCTGCGAGGCCCTGGGCTTCAACACAGAGTCCCGCCAGCAGCC 2632
3357 CATGGCCCTGGCTGTGGCCCTAACCAAGGTGGAGAGGCCCGAGGGGAGCTGTTCGGGA 3416
2633 CATGGCCCTGGCTGTGGCCCTGACCAAGGTGGGAGGCCCGAGGGGAGCTGTTCGGGA 2692
3417 CGATGGAGAGAGCTGGAAAGTCTGAGCGAGGGGCTTACACAGAGTCAATCTTCCTGCG 3476
2693 CGATGGAGAGAGCTGGAAAGTCTGAGCGAGGGGCTTACACAGAGTCAATCTTCCTGCG 2752
3477 CAGGAATACAGATCGTGAATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3536
2753 CAGGAATACAGATCGTGAATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2812
3537 GCTGACAGAGGTGAGTGTCTGGGGGTGGCCACGGCGCCAGCAGAGTCTCTTCCAAACGG 3596
2813 GCTGACAGAGGTGAGTGTCTGGGGGTGGCCACGGCGCCAGCAGAGTCTCTTCCAAACGG 2872
3597 TGTCCCTGTCTCCAACTTACCTACAGCGCCCGACACCAAGGTCCTGGAGCATCTGTGTCTC 3656
2873 TGTCCCTGTCTCCAACTTACCTACAGCGCCCGACACCAAGGTCCTGGAGCATCTGTGTCTC 2932
3657 GCTGTGTGTGGAGAGCAGTTTCTGTCAGCTGGTGT 3694
2933 GCTGTGTGTGGAGAGCAGTTTCTGTCAGCTGGTGT 2970

DT	18-NOV-2004	(first entry)	
XX	Human diagnostic and therapeutic polynucleotide SEQ ID NO:1844.		
XX	ss: gene; gene therapy; human diagnostic and therapeutic polynucleotide;		
KW	dithp.		
XX	Homo sapiens.		
OS	WO2004023973-A2.		
XX	25-MAR-2004.		
XX	12-SEP-2003; 2003WO-US028227.		
XX	12-SEP-2002; 2002US-0410259P.		
PR	12-SEP-2002; 2002US-0410260P.		
XX	(INCY-) INCYTE CORP.		
XX	Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F, Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV; Mooney EM, Delegane AM, Panesar IS, Banville SC, Reddy RP; Stevens KA, Blanchard JL, Rioux P, Wang X, Au AP, Gerstein EH; Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL; Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtan ES; Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D; Patury S, Shi X, Suarez CJ;		
XX	WPI; 2004-329368/30.		
DR	P-PSDB; ABM84317.		
XX	New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.		
XX	Claim 1; Page; 190pp; English.		
XX	The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2722 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, autoimmune/inflammatory disorder, developmental disorder, endocrine disorder, neurological disorders, gastrointestinal disorders, or infections caused by virus, bacteria, fungi or parasite. The dithp molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline gene therapy. The present sequence represents a dithp polynucleotide of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm		
XX	Sequence 2655 BP; 455 A; 900 C; 814 G; 486 T; 0 U; 0 Other;		
SQ	Query Match 32.2%; Score 1191.6; DB 13; Length 2655; Best Local Similarity 93.6%; Pred. No. 7.3e-214; Matches 1286; Conservative 0; Mismatches 9; Indels 79; Gaps 1;		
QY	1036	GAACCGGTGCACACCCCGCGTCCAGAGCAGTGCCACACAGTGCAGCGTCCCGCC	1095
DB	454	GATGCCAGGACACCCCGCGCGTCCAGAGCAGTGCCACACAGTGCAGCGTCCCGCC	513
QY	1096	AACAGCGGTTCGATTGCGCCCTGCAAGGCCATCACCCAGGAACAGTGCAGGCGCCG	1155
DB	514	AACAGCGGTTCGATTGCGCCCTGACAGGCCATCACCCAGGAACAGTGCAGGCGCCG	573
QY	1156	GGCTGTGTACATCCCTGCAAGCAGGGGTGTCAGGAGGCCAGATGGGGGAGCCCTGG	1215
DB	574	GGCTGTGTACATCCCTGCAAGCAGGGGTGTCAGGAGGCCAGATGGGGGAGCCCTGG	633
QY	1216	TCCTTCTTCCCAACCAGCTACCCCAAGCTACAAGCTGGAGAACCTGAGTCTCTTGAATG	1275
DB	634	TGCTTCTTCCCAACCAGCTACCCCAAGCTACAAGCTGGAGAACCTGAGTCTCTTGAATG	693
QY	1276	GGCTACAGGGCCACCTGACCGGTACCAACCCCAAGCTTCTTCCCAAGGACATCCTGACC	1335
DB	694	GGCTACAGGGCCACCTGACCGGTACCAACCCCAAGCTTCTTCCCAAGGACATCCTGACC	753
QY	1336	CTGCGGCTGGAGCTGATGATGAGAGCTGAGAACCCGCTCCACTTTCACGATCAAGATCCA	1395
DB	754	CTGCGGCTGGAGCTGATGATGAGAGCTGAGAACCCGCTCCACTTTCACGATCAAGATCCA	813
QY	1396	GCTAACAGCGCTACGAGTGCCTTGGAGAGACCCCGCTGTCCACAGCGGACCCGCTCC	1455
DB	814	GCTAACAGCGCTACGAGTGCCTTGGAGAGACCCCGCTGTCCACAGCGGACCCGCTCC	873
QY	1456	CCACTCTACAGCGTGGAGTTCCTCGAGGAGCCCTTCCGGGTGATCGTGGCGGCGAGCTG	1515
DB	874	CCACTCTACAGCGTGGAGTTCCTCGAGGAGCCCTTCCGGGTGATCGTGGCGGCGAGCTG	933
QY	1516	GACGCGCGGTGCTGCTGAACACGAGCGGTGGCGCCCTGTTCTTTGCGGACCAAGTTCCTT	1575
DB	934	GACGCGCGGTGCTGCTGNAACAGAGCGGTGGCGCCCTGTTCTTTGCGGACCAAGTTCCTT	993
QY	1576	CAGCTGTCCACCTCGCTGCGCTGCAGTATATACAGAGGCTTCGCGAGACCTCAGTCCC	1635
DB	994	CAGCTGTCCACCTCGCTGCGCTGCAGTATATACAGAGGCTTCGCGAGACCTCAGTCCC	1053
QY	1636	CTGATGCTCAGCACAGCTGGACAGGATCACCCCTGTGGAACCGGACCTTGGCGCCACG	1695
DB	1054	CTGATGCTCAGCACAGCTGGACAGGATCACCCCTGTGGAACCGGACCTTGGCGCCACG	1113
QY	1696	CCCGGTGGAGACCTCTACGGGTCTCACCCCTTCTACCTGGCGCTGGAGACCGCGGTG	1755
DB	1114	CCCGGTGGAGACCTCTACGGGTCTCACCCCTTCTACCTGGCGCTGGAGACCGCGGTG	1173
QY	1756	GCACACGGGGTTCCTGCTAAACAGCAATGCATGGATGTGCTCTGAGCGGACCCCT	1815
DB	1174	GCACACGGGGTTCCTGCTAAACAGCAATGCATGGATGTGCTCTGAGCGGACCCCT	1233
QY	1816	GGCTTAGCTGAGGTGCAGGTGGATCTCTGGATGTACATCTTCTGGGCGCCAGAG	1875
DB	1234	GGCTTAGCTGAGGTGCAGGTGGATCTCTGGATGTACATCTTCTGGGCGCCAGAG	1293
QY	1876	CCCAAGAGCGTGTGTCAGCAGTACCTGACGTGTG-----	1911
DB	1294	CCCAAGAGCGTGTGTCAGCAGTACCTGACGTGTG-----	1353
QY	1912	-----GGATA	1916
DB	1354	GCCCCCGCCCCAAGGCTCCCTCCCTCCCTCATGAAGTCGGCGTGGCTGCAGGATA	1413
QY	1917	CCCGTTATGCGCGCATACTGGGGCTGGGCTTCCACCTGTGCGGTGGGGCTACTCTTC	1976
DB	1414	CCCGTTATGCGCGCATACTGGGGCTGGGCTTCCACCTGTGCGGTGGGGCTACTCTTC	1473
QY	1977	CACCGCTATACCCCGCAGGTGTGAGAACATGACAGGGCCCACTTCCCTCGACGT	2036
DB	1474	CACCGCTATACCCCGCAGGTGTGAGAACATGACAGGGCCCACTTCCCTCGACGT	1533
QY	2037	CCAATGGAACGACCTGGAGTACATGGACTTCCCGGAGGAGCTTTCACGTTCAAGAGATG	2096
DB	1534	CCAATGGAACGACCTGGAGTACATGGACTTCCCGGAGGAGCTTTCACGTTCAAGAGATG	1593
QY	2097	CTTCCGGAGCTTCCCGGCGATGGTGCAGAGTGCACAGGGGCGCGCGCTCATGAT	2156
DB	1594	CTTCCGGAGCTTCCCGGCGATGGTGCAGAGTGCACAGGGGCGCGCGCTCATGAT	1653
QY	2157	GATCGTGGATCTGCGCATCAGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTACGACCA	2216
DB	1654	GATCGTGGATCTGCGCATCAGCAGCTCGGGCCCTCGCGGAGCTACAGGCCCTACGACCA	1713

QY 2217 GGGTCTGCGAGGGGGTTCATCAACAAACGAGACCGCCAGCCGCTGATTGGGAAGGT 2276
DB 1714 GGGTCTGCGAGGGGGTTCATCAACAAACGAGACCGCCAGCCGCTGATTGGGAAGGT 1773
QY 2277 ATGGCCCGGGTCCACTGCTTCCCGCACTTCCACCAACCCACAGCCCTGGCCTG 2330
DB 1774 ATGGCCCGGGTCCACTGCTTCCCGCACTTCCACCAACCCACAGCCCTGGCCTG 1827

RESULT 11

AAC87527

ID AAC87527 standard; DNA; 1493 BP.

XX

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

AC

Human RAP (receptor associated protein) cDNA.

RAP; receptor associated protein; human; factor VIII clearance;

LRP-mediated plasma clearance; receptor-dependent clearance;

receptor-independent clearance; ligand internalisation;

low density lipoprotein related protein; haemophilia; half-life; ss.

Homo sapiens.

WO200071714-A2.

30-NOV-2000.

24-MAY-2000; 2000WO-US014111.

24-MAY-1999; 99US-0135847P.

(AMNA-) AMERICAN NAT RED CROSS.

Saenko EL, Strickland DK;

WPI; 2001-025163/03.

P-PSDB; AAB48844.

Factor VIII mutants having increased half-life useful for treating

hemophilia, comprise one or more amino acid substitutions in the A2

and/or C2 domain of factor VIII.

Disclosure; Page 111-114; 121pp; English.

The invention relates to human factor VIII mutants comprising an amino

acid substitution at one or more positions in the A2 domain and/or an

amino acid substitution at one or more positions in the C2 domain. The

invention also encompasses a factor VIII mutant which lacks a B domain

(AAB48842). The factor VIII mutants have an increased half-life in the

bloodstream. The A2 domain mutants exhibit reduced LRP-dependent

(receptor-dependent) clearance of factor VIII, while C2 domain mutants

have reduced receptor-independent clearance. The invention also relates

to a method of using RAP (receptor associated protein), a protein which

inhibits LRP (low density lipoprotein related protein)-mediated ligand

internalisation, to increase the half-life of factor VIII. The mutant

factor VIII proteins, and nucleotides encoding them, are useful for

treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and

nucleic acids encoding them may also be used in the treatment of

haemophilia. In combination with a mutant factor VIII protein or DNA of

the invention. The invention provides means of increasing the half-life

of factor VIII by reducing its clearance from plasma. The present

sequence represents cDNA encoding human RAP

Sequence 1493 BP; 360 A; 412 C; 488 G; 233 T; 0 U; 0 Other;

Query Match 25.9%; Score 957; DB 5; Length 1493;

Best Local Similarity 99.0%; Pred. No. 6.7e-170;

Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCCTACTCGCGGAGAGAAACGAGCCGTCGCCGAAACGGCGAGTCCGGAGA 122

DB 1109 CGGCAAGTACTCGCGGAGAGAAACAGCCCAAGCCGTCGCCGAAACGCGAGTCCGGAGA 168
QY 123 GGAGTTCGCGATGGAGAAAGTTGAACACAGCTGTGGGAGAGAGCCAGCGACTGCATCTTCC 182
DB 169 GGAGTTCGCGATGGAGAAAGTTGAACACAGCTGTGGGAGAGAGCCAGCGACTGCATCTTCC 228
QY 183 TCCCGTGGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGAGAGAGGAGCGAACTCGC 242
DB 229 TCCCGTGGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGAGAGAGGAGCGAACTCGC 288
QY 243 CTGGAAGAAACTAAAGCTTGAAGCTTGGACGAAAGATCGGAGAGAGAGAGAGAGAGAGAGAG 302
DB 289 CTGGAAGAAACTAAAGCTTGAAGCTTGGACGAAAGATCGGAGAGAGAGAGAGAGAGAGAG 348
QY 303 ACGCAACCTCAATGTCTTGGCCCAAGTATGCTGTGGACGGAAGAGAGAGAGAGAGAGAGAG 362
DB 349 ACGCAACCTCAATGTCTTGGCCCAAGTATGCTGTGGACGGAAGAGAGAGAGAGAGAGAGAG 408
QY 363 GGTGACCAAGCAACTCCCTCAGTGGCACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422
DB 409 GGTGACCAAGCAACTCCCTCAGTGGCACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 468
QY 423 AAAGCTGTGGCAACAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAGAGAGAGAGAGAG 482
DB 469 AAAGCTGTGGCAACAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAGAGAGAGAGAG 528
QY 483 CTGGCGGAGGTTCTCTGCATCAAAAGAGAAAGTTCAACAGATACAAAGAGAGAGAGAGAGAG 542
DB 529 CTGGCGGAGGTTCTCTGCATCAAAAGAGAAAGTTCAACAGATACAAAGAGAGAGAGAGAG 588
QY 543 CCTGAGCAGGAGCCGAGAAATTCACAGAGAACTCAATAGCCCTCCGAGAGAGAGAGAGAGAG 602
DB 589 CCTGAGCAGGAGCCGAGAAATTCACAGAGAACTCAATAGCCCTCCGAGAGAGAGAGAGAGAG 648
QY 603 CAAAGGCGAGGCTCTGCAACAG 662
DB 649 CAAAGGCGAGGCTCTGCAACAG 708
QY 663 CCAGGCGCTGGACCGCCCTGCGCAGGCTCAGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 722
DB 709 CCAGGCGCTGGACCGCCCTGCGCAGGCTCAGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 768
QY 723 CGAGGAGCCAGGAGTGTGACCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782
DB 769 CGAGGAGCCAGGAGTGTGACCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 828
QY 783 GGAGCTGGAGGCTTCCGCGAGGAGCTCAGCACTTCGAGCCAAATTCGAGAGAGAGAGAGAGAG 842
DB 829 GGAGCTGGAGGCTTCCGCGAGGAGCTCAGCACTTCGAGCCAAATTCGAGAGAGAGAGAGAGAG 888
QY 843 CCAGCTACAG 902
DB 889 CCAGCTACAG 948
QY 903 CGAGCGGAGGCTGTGAGCCGCGAG 962
DB 949 CGAGCGGAGGCTGTGAGCCGCGAG 1008
QY 963 GGAGCTGGGCTACAGGCTGAG 1022
DB 1009 GGAGCTGGGCTACAGGCTGAG 1068
QY 1023 TCGCGCCGAGGCA 1035
DB 1069 TCGGCACAAACGAA 1081

RESULT 12

AAL48893

ID AAL48893 standard; cDNA; 1493 BP.

XX

AC AAL48893;

gene therapy; marker gene; gene; ds.
 Homo sapiens.
 EP1394274-A2.
 03-MAR-2004.
 04-AUG-2003; 2003BP-00254857.
 06-AUG-2002; 2002JP-00229312.
 20-MAR-2003; 2003JP-00077212.
 (GENO-) GENOX RES INC.
 Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuwara K;
 WPI; 2004-193155/19.
 Testing for bronchial asthma or chronic obstructive pulmonary disease by
 comparing the expression level of a marker gene in a biological sample
 from a subject with the expression level of the gene in a sample from a
 healthy subject.
 Claim 1; SEQ ID NO 92; 241pp; English.
 The present invention describes a method of testing for bronchial asthma
 or chronic obstructive pulmonary disease. The method comprises
 determining the expression level of a marker gene in a biological sample
 from a subject, comparing the expression level determined with the
 expression level of the marker gene in a biological sample from a healthy
 subject, and judging whether the subject has bronchial asthma or chronic
 obstructive pulmonary disease. The marker gene comprises: (a) a group of
 genes (S1) whose expression levels increase when respiratory epithelial
 cells are stimulated with interleukin-13; or (b) a group of genes (S2)
 whose expression levels decrease when respiratory epithelial cells are
 stimulated with interleukin-13. Also described: (1) a reagent (I) for
 testing for bronchial asthma or chronic obstructive pulmonary disease;
 (2) a kit for screening for a candidate compound for a therapeutic agent
 to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
 an animal model for bronchial asthma or chronic obstructive pulmonary
 disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
 method for producing an animal model for bronchial asthma or chronic
 obstructive pulmonary disease; (6) a therapeutic agent for bronchial
 asthma or chronic obstructive pulmonary disease, comprising the compound,
 a marker gene or an antisense nucleic acid corresponding to a portion of
 the marker gene, a ribozyme, a polynucleotide that suppresses the
 expression of the gene through an RNAi effect or an antibody recognising
 a protein encoded by a marker gene; and (7) a DNA chip for testing for
 bronchial asthma or a chronic obstructive pulmonary disease, on which a
 probe has been immobilised to assay a marker gene. (I) has respiratory
 and antiasthmatic activities, and can be used in gene therapy. The method
 is useful for testing for or screening for a therapeutic agent for
 bronchial asthma or chronic obstructive pulmonary disease. The present
 sequence is used in the exemplification of the present invention.
 Sequence 1493 BP; 360 A; 412 C; 488 G; 233 T; 0 U; 0 Other;
 Query Match 25.9%; Score 957; DB 12; Length 1493;
 Best Local Similarity 99.0%; Pred. No. 6.7e-170;
 Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 63 CGGATCCTACTCGCGGAGAGAACAGCCCGTCCCGAAACGGAGTCCGGAGA 122
 DB 109 CGGCAAGTACTCGCGGAGAGAGAACAGCCCGTCCCGAAACGGAGTCCGGAGA 168
 QY 123 GGAGTTCCGATGGAGAGTTGAACAGCTGTGGGAGAGGCCAGCGACTGCATCTTCC 182
 DB 169 GGAGTTCCGATGGAGAGTTGAACAGCTGTGGGAGAGGCCAGCGACTGCATCTTCC 228
 QY 183 TCCCGTGGAGTGGCGGAGTCCAGCTGTGAAGATACAGGAGGAGCGAACTCCG 242
 DB 229 TCCCGTGGAGTGGCGGAGTCCAGCTGTGAAGATACAGGAGGAGCGAACTCCG 288

QY 243 CTGGAAGAACTAAAGCTTTGACGGCTTGACGAAAGATGGGGAGAGAAAGCGAGACTCAT 302
 DB 289 CTGGAAGAACTAAAGCTTTGACGGCTTGACGAAAGATGGGGAGAGAAAGCGAGACTCAT 348
 QY 303 ACGCAACCTCAATGTCTATCTTGGCAAGATATGGTCTGGACCGAAAGAAAGCGCTCGGCA 362
 DB 349 ACGCAACCTCAATGTCTATCTTGGCCAGATATGGTCTGGACCGAAAGAAAGCGCTCGGCA 408
 QY 363 GGTGACCAAGCAATCCCTCAGTGGCAACCAGGAAGACGGGCTGGATGACCCCAAGGTGGA 422
 DB 409 GGTGACCAAGCAATCCCTCAGTGGCAACCAGGAAGACGGGCTGGATGACCCCAAGGTGGA 468
 QY 423 AAAGCTGTGGCAAGCGAAGACCTCTCGGAAAATTTCTCGGCGAAGAACTGACAAAGCT 482
 DB 469 AAAGCTGTGGCAAGCGAAGACCTCTCGGAAAATTTCTCGGCGAAGAACTGACAAAGCT 528
 QY 483 CTGGCGGAGTTCTTGCATCACAAAGAGAAAGTTTACAGAGTACAAAGTCTCTGCTGGAGAC 542
 DB 529 CTGGCGGAGTTCTTGCATCACAAAGAGAAAGTTTACAGAGTACAAAGTCTCTGCTGGAGAC 588
 QY 543 CCTGAGCAGGACCGGAAGAAATCCACGAGAACGTCATTAGCCCCCTCGGACCTTGAGCGACAT 602
 DB 589 CCTGAGCAGGACCGGAAGAAATCCACGAGAACGTCATTAGCCCCCTCGGACCTTGAGCGACAT 648
 QY 603 CAAGGCGAGCGTCTTGCACAGCAGGACACACGAGCTGAAAGAGAAAGCTGCGCAGCATCAA 662
 DB 649 CAAGGCGAGCGTCTTGCACAGCAGGACACACGAGCTGAAAGAGAAAGCTGCGCAGCATCAA 708
 QY 663 CCAGGCGTGGACCGCTTGCAGGCTCAGCCAGGCTCAGCCAGGCTCAGCAGCTGAGGCTGAGTT 722
 DB 709 CCAGGCGTGGACCGCTTGCAGGCTCAGCCAGGCTCAGCCAGGCTCAGCAGCTGAGGCTGAGTT 768
 QY 723 CGAGGAGCCAGCGGTGATTGACCTGTGGGACCTTGGCGAGTCCGCCAACTCTCACGGACAA 782
 DB 769 CGAGGAGCCAGCGGTGATTGACCTGTGGGACCTTGGCGAGTCCGCCAACTCTCACGGACAA 828
 QY 783 GGAGCTGGAGGCGTTCGGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAAGCACAA 842
 DB 829 GGAGCTGGAGGCGTTCGGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAAGCACAA 888
 QY 843 CCAGTACAGAGCAGCTGAGATTGCGCACGAGAGCTGAGCGACGCGAGAGCGCTGGG 902
 DB 889 CCAGTACAGAGCAGCTGAGATTGCGCACGAGAGCTGAGCGACGCGAGAGCGCTGGG 948
 QY 903 CGACGCGAGCGTGTGAGCGCGAGCGCGAGAGCAAGCAAGCGCTGCTGGAGGGCGGACCAA 962
 DB 949 CGACGCGAGCGTGTGAGCGCGAGCGCGAGAGCAAGCAAGCGCTGCTGGAGGGCGGACCAA 1008
 QY 963 GGAGCTGGGCTACCGGTGAAAGAGCATGTGAGGACCTGTCCGGCAGGATTTCCAGAGC 1022
 DB 1009 GGAGCTGGGCTACCGGTGAAAGAGCATGTGAGGACCTGTCCGGCAGGATTTCCAGAGC 1068
 QY 1023 TCGCGCGGAGGCA 1035
 DB 1069 TCGGCACACGAA 1081

RESULT 15

ADR24896

ID ADR24896 standard; DNA; 1493 BP.

XX ADR24896;

DT 21-OCT-2004 (first entry)

DE Breast cancer prognosis marker #757.

XX ds; breast cancer; prognosis; gene expression; diagnosis.

XX Homo sapiens.

PN WO2004065545-A2.

XX PD 05-AUG-2004.
XX XX
PF 15-JAN-2004; 2004WO-US001100.
XX PR 15-JAN-2003; 2003US-00342887.
XX PA (ROSE-) ROSETTA INPHARMATICS LLC.
XX PA (NECA-) NETHERLANDS CANCER INST.
XX PI Van't Veer LJ, He Y;
XX WPI; 2004-593473/57.
XX PT Classifying a breast cancer patient according to prognosis comprises
XX PT determining the similarity between the level of expression of each of
XX PT five genes in a cell sample taken from patient, to control levels.
XX PS Disclosure; SEQ ID NO 757; 226pp; English.
XX CC The invention relates to a method of classifying a breast cancer patient
XX CC according to prognosis by determining the similarity between the level of
XX CC expression of each of five genes for which markers are listed in the
XX CC specification, in a cell sample taken from the breast cancer patient, to
XX CC control levels of expression for each respective five genes to obtain a
XX CC patient similarity value. The methods are useful for classifying a breast
XX CC cancer patient according to prognosis. Kits and computer program products
XX CC are useful for data analysis using the diagnostic, prognostic and
XX CC statistical methods of the invention. This sequence corresponds to a
XX CC marker used in the method of the invention.
XX SQ Sequence 1493 BP; 360 A; 412 C; 488 G; 233 T; 0 U; 0 Other;

Query Match 25.9%; Score 957; DB 13; Length 1493;
Best Local Similarity 99.0%; Pred. No. 6.7e-170;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCCTACTCGCGGAGAAACAGCCAGCGCTCCCGAAACGGAGTCCGGAGA 122
DB |||||
QY 123 GGAGTCCCGATGGAGAAAGTTGAACAGCTGTGGGAGAAAGCCAGCGCTGCATCTCC 182
DB |||||
QY 169 GGAGTCCCGATGGAGAAAGTTGAACAGCTGTGGGAGAAAGCCAGCGCTGCATCTCC 228
QY 183 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAGATACAGGAGAGCGAACTCGC 242
DB |||||
QY 229 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAACTCGC 288
QY 243 CTGGAGAAAGCTAAAGCTTTGACGGCTTGGACGAAGATGGGAGAGAAAGCGAGACTCAT 302
DB |||||
QY 289 CTGGAGAAAGCTAAAGCTTTGACGGCTTGGACGAAGATGGGAGAGAAAGCGAGACTCAT 348
QY 303 ACGCAACCTCAATGTCTATCTTGCCCAAGTATGGTCTGGACGGAAGAAAGAGCGCTCGGCA 362
DB |||||
QY 349 ACGCAACCTCAATGTCTATCTTGCCCAAGTATGGTCTGGACGGAAGAAAGAGCGCTCGGCA 408
QY 363 GGTGACCAAGCTCCCTCAGTGGGACCCAGGAGAGCGGCTGGATGACCCAGGCTGGA 422
DB |||||
QY 409 GGTGACCAAGCTCCCTCAGTGGGACCCAGGAGAGCGGCTGGATGACCCAGGCTGGA 468
QY 423 AAAGCTGTGGCACAAGCGGAAGCTCTTGGGAAATTTCTCCGGCGAAGAACTGGACAAGCT 482
DB |||||
QY 469 AAAGCTGTGGCACAAGCGGAAGCTCTTGGGAAATTTCTCCGGCGAAGAACTGGACAAGCT 528
QY 483 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTCAAGAGTCAACAGTCTCTGTGGAGAC 542
DB |||||
QY 529 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTCAAGAGTCAACAGTCTCTGTGGAGAC 588
QY 543 CTTGAGCAGGACCGGAGAAATCCACAGAGAAAGCTCATTAGCCCTCGGACCTGAGGACAT 602
DB |||||
QY 589 CTTGAGCAGGACCGGAGAAATCCACAGAGAAAGCTCATTAGCCCTCGGACCTGAGGACAT 648

Search completed: April 4, 2005, 11:37:46
JOB time : 1774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2005, 11:07:54 ; Search time 608 Seconds
(without alignments)
9962.988 Million cell updates/sec

Title: US-10-600-862A-6
Perfect score: 3702
Sequence: 1 cttaccgccatcg99gtcc.....tcagctggtgtgactcgag 3702

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2647.8	71.5	3846	4 US-09-949-016-23	Sequence 23, Appl
2	2644.6	71.4	3847	4 US-09-949-016-3656	Sequence 3656, Ap
3	370.6	10.0	449	3 US-09-397-787-175	Sequence 175, App
4	369.8	10.0	2752	1 US-08-430-925A-3	Sequence 3, Appli
5	340.4	9.2	22404	4 US-09-949-016-11765	Sequence 11765, A
6	340.4	9.2	22404	4 US-09-949-016-15398	Sequence 15398, A
7	212.2	5.7	2272	4 US-09-744-926-1	Sequence 1, Appli
8	189.8	5.1	3018	4 US-09-949-016-4756	Sequence 4756, Ap
9	175.2	4.7	601	4 US-09-949-016-18094	Sequence 18094, A
10	175.2	4.7	601	4 US-09-949-016-128902	Sequence 128902,
11	165.2	4.5	601	4 US-09-949-016-18093	Sequence 18093, A
12	165.2	4.5	601	4 US-09-949-016-128901	Sequence 128901,
13	157.4	4.3	601	4 US-09-949-016-18081	Sequence 18081, A
14	157.4	4.3	601	4 US-09-949-016-128889	Sequence 128889,
15	155	4.2	601	4 US-09-949-016-18077	Sequence 18077, A
16	155	4.2	601	4 US-09-949-016-128885	Sequence 128885,
17	153.8	4.2	601	4 US-09-949-016-18075	Sequence 18075, A
18	153.8	4.2	601	4 US-09-949-016-18076	Sequence 18076, A
19	153.8	4.2	601	4 US-09-949-016-128883	Sequence 128883,
20	153.8	4.2	601	4 US-09-949-016-128884	Sequence 128884,
21	153.4	4.1	601	4 US-09-949-016-18078	Sequence 18078, A
22	153.4	4.1	601	4 US-09-949-016-18079	Sequence 18079, A
23	153.4	4.1	601	4 US-09-949-016-128886	Sequence 128886,
24	153.4	4.1	601	4 US-09-949-016-128887	Sequence 128887,
25	149.6	4.0	601	4 US-09-949-016-18080	Sequence 18080, A
26	149.6	4.0	601	4 US-09-949-016-128888	Sequence 128888,
27	135.8	3.7	601	4 US-09-949-016-18082	Sequence 18082, A

ALIGNMENTS

RESULT 1

US-09-949-016-23
; Sequence 23, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 3846
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-23

Query Match	71.5%	Score	2647.8;	DB	4;	Length	3846;
Best Local Similarity	99.7%	Pred.	No. 0;				
Matches	2652;	Conservative	0;	Mismatches	7;	Indels	0;
Gaps	0;						
QY	1036	GAACCGGTGCACACCCCGCGTCCACAGAGTGTCCACAGAGTGTCCACAGTGGCAGTGCCTCCCTCC	1095				
DB	640	GATGCCAGGACACACCCCGCGTCCACAGAGTGTCCACAGAGTGTCCACAGTGGCAGTGCCTCCCTCC	699				
QY	1096	AACAGCGGTTCGATTGCGCCCTTGACAGAGGCGCATCACCAGGAAACAGTGCAGAGGCGCGC	1155				
DB	700	AACAGCGGTTCGATTGCGCCCTTGACAGAGGCGCATCACCAGGAAACAGTGCAGAGGCGCGC	759				
QY	1156	GGCTGTGTACATCCCTGCAAGCAGGGGCTGCAGGGAGCCAGATGGGGAGCCCTGG	1215				
DB	760	GGCTGTGTACATCCCTGCAAGCAGGGGCTGCAGGGAGCCAGATGGGGAGCCCTGG	819				
QY	1216	TGCTTTTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTTGAATG	1275				
DB	820	TGCTTTTCCACCCAGCTACCCAGCTACAGCTGGAGAACCTGAGTCTCTTGAATG	879				
QY	1276	GGCTACAGCGCCACCTGACCCCTGACCCACCCCTCTTCTCCCAAGGACATCTCTGACC	1335				
DB	880	GGCTACAGCGCCACCTGACCCCTGACCCACCCCTCTTCTCCCAAGGACATCTCTGACC	939				
QY	1336	CTGCGGCTGGAGCTGATGATGGAGACTGAGAACCCGCTCCACTTCAGATCAAGATCCA	1395				

Db	940		CTGCGCTGAGCGTATGATGAGACTGAGAACCGCTCCATCTTACGATCAAGATCCA	999
QY	1396		GCTTAACAGCGCTTACAGGTCCTTGGAGACCCCGCGTGTCCACAGCGGGCACCGTCC	1455
Db	1000		GCTTAACAGCGCTTACAGGTCCTTGGAGACCCCGCGTGTCCACAGCGGGCACCGTCC	1059
QY	1456		CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGTATCGTGACCGGACAGCTG	1515
Db	1060		CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGTATCGTGACCGGACAGCTG	1119
QY	1516		GACGGCGGTGTGTCTGAACAGAGCGGTGGCGCCCTGTCTTCTTTCGGGACCGATTCCTT	1575
Db	1120		GACGGCGGTGTGTCTGAACAGAGCGGTGGCGCCCTGTCTTCTTTCGGGACCGATTCCTT	1179
QY	1576		CAGCTGTCCACCTCGCTGCGCTGCGAGTATATCACAGGCTCGCCAGCACCTCAGTCCC	1635
Db	1180		CAGCTGTCCACCTCGCTGCGCTGCGAGTATATCACAGGCTCGCCAGCACCTCAGTCCC	1239
QY	1636		CTGATGCTCAGCACAGCTGGACAGGATCACCTGTGGAAACCGGACCTTGGCCCAAG	1695
Db	1240		CTGATGCTCAGCACAGCTGGACAGGATCACCTGTGGAAACCGGACCTTGGCCCAAG	1299
QY	1696		CCCGGTGCGAACTCTACGCGTCTACCCCTTCTACTCTGCGGTGGAGACGGCGGGTCG	1755
Db	1300		CCCGGTGCGAACTCTACGCGTCTACCCCTTCTACTCTGCGGTGGAGACGGCGGGTCG	1359
QY	1756		GCACAGGGGTGTTCTGTCTTAACAGCAATGCCATGGATGTGTCTGCGAGCGAGCCCT	1815
Db	1360		GCACAGGGGTGTTCTGTCTTAACAGCAATGCCATGGATGTGTCTGCGAGCGAGCCCT	1419
QY	1816		GCCCTTAGCTGGAGTGCAGAGTGGGATCTTGGATGTCTACTTCTTGGGCCCAAGAG	1875
Db	1420		GCCCTTAGCTGGAGTGCAGAGTGGGATCTTGGATGTCTACTTCTTGGGCCCAAGAG	1479
QY	1876		CCCAAGAGCTGTGTGAGAGTACCTTGGAGTGTGTGGGATACCCGTTCATGCGGCCATAC	1935
Db	1480		CCCAAGAGCTGTGTGAGAGTACCTTGGAGTGTGTGGGATACCCGTTCATGCGGCCATAC	1539
QY	1936		TGGGGCTTGGCTTCCACCTGTGCGCTGGGGTACTCTCCACCGCTATACCCGCCAG	1995
Db	1540		TGGGGCTTGGCTTCCACCTGTGCGCTGGGGTACTCTCTCCACCGCTATACCCGCCAG	1599
QY	1996		GTGTTGGAGAACATGACAGGGGCCACTTCCCCCTTGGAGCTCCAAATGGAAACGACTGGAC	2055
Db	1600		GTGTTGGAGAACATGACAGGGGCCACTTCCCCCTTGGAGCTCCAAATGGAAACGACTGGAC	1659
QY	2056		TACATGGAATCCCGAGGGACTTTCAGTTCAACAGGATGGCTTCCGGGACTTCCCGGCC	2115
Db	1660		TACATGGAATCCCGAGGGACTTTCAGTTCAACAGGATGGCTTCCCGGACTTCCCGGCC	1719
QY	2116		ATGGTCAGAGCTGCACAGGGGCGCGCGCTACATGATGATCGTGGATCCTGCCATC	2175
Db	1720		ATGGTCAGAGCTGCACAGGGGCGCGCGCTACATGATGATCGTGGATCCTGCCATC	1779
QY	2176		AGCAGCTCGGGCTCTCGGGAGCTACAGGCCCTTACGAGGGTCTCGGAGGGGGTT	2235
Db	1780		AGCAGCTCGGGCTCTCGGGAGCTACAGGCCCTTACGAGGGTCTCGGAGGGGGTT	1839
QY	2236		TTTATCAACAGAGCCCGGCTGATTTGGGAAAGTATGGCCCGGGTTCATGCGC	2295
Db	1840		TTTATCAACAGAGCCCGGCTGATTTGGGAAAGTATGGCCCGGGTTCATGCGC	1899
QY	2296		TTTCCCGACTTCAACCAACCCAGCCCTGCTGTGGAGGACATGGTGGTGGATTC	2355
Db	1900		TTTCCCGACTTCAACCAACCCAGCCCTGCTGTGGAGGACATGGTGGTGGATTC	1959
QY	2356		CATGACAGGTGCTTTCGAGCGCTTGTGATGATGATGAACAGAGCTTCCAACTTCATC	2415
Db	1960		CATGACAGGTGCTTTCGAGCGCATGTGATGATGATGAACAGAGCTTCCAACTTCATC	2019
QY	2416		AGAGGCTGTGAGAGCGGCTGCCCAACATAGCTGGAGAACCCACCTTACGTGCTGGG	2475

Db	2020		AGAGGCTGTGAGAGCGGCTGCGGCAACCAATAGCTGGAGAAACCCACCTTACGTGCTGGG	2079
QY	2476		GTGGTTGGGGGACCTTCCAGGCGCCACCAATCTGTGCTCCAGGACACCAAGTTTCTCTCC	2535
Db	2080		GTGGTTGGGGGACCTTCCAGGCGCCACCAATCTGTGCTCCAGGACACCAAGTTTCTCTCC	2139
QY	2536		ACACACTACAACTCTGACAACTCTTACGGCTTGAACGAAGCCATCGCTCCACAGGGCG	2595
Db	2140		ACACACTACAACTCTGACAACTCTTACGGCTTGAACGAAGCCATCGCTCCACAGGGCG	2199
QY	2596		CTGGTGAAGGCTCGGGGACACCGCCATTTGTGATCTCCCGCTCGACCTTTCTGCTGCGCAC	2655
Db	2200		CTGGTGAAGGCTCGGGGACACCGCCATTTGTGATCTCCCGCTCGACCTTTCTGCTGCGCAC	2259
QY	2656		GGCCATATACCGCGGCACTTGACGGGGACGTGTGGAGCTCTTGGGAGCAGCTCGCTCC	2715
Db	2260		GGCCATATACCGCGGCACTTGACGGGGACGTGTGGAGCTCTTGGGAGCAGCTCGCTCC	2319
QY	2716		TCCTGTCAGAAATCTTGCAGTTTAACTGTGTGGGGTGCCTCTGTGTGGGGCCGACGTC	2775
Db	2320		TCCTGTCAGAAATCTTGCAGTTTAACTGTGTGGGGTGCCTCTGTGTGGGGCCGACGTC	2379
QY	2776		TGCGGCTTCTGGGCAACACTCAGAGGAGCTGTGTGGGCTGACCCAGCTGGGGCC	2835
Db	2380		TGCGGCTTCTGGGCAACACTCAGAGGAGCTGTGTGGGCTGACCCAGCTGGGGCC	2439
QY	2836		TTTACTACCTTTCATGCGGAAACCAACAGACCTGTCTCAGTCTGCCCCAGAGCGCTACAGC	2895
Db	2440		TTTACTACCTTTCATGCGGAAACCAACAGACCTGTCTCAGTCTGCCCCAGAGCGCTACAGC	2499
QY	2896		TTTACGAGAGCGGCGCAGAGCCATGAGGAAGGCTCTACCTTGGCTTACGACCTCTCTC	2955
Db	2500		TTTACGAGAGCGGCGCAGAGCCATGAGGAAGGCTCTACCTTGGCTTACGACCTCTCTC	2559
QY	2956		CCCACTCTACACTGTTTCCACAGGCGCCAGCTGCGGGGGAGACCGTGGCCCGGGCC	3015
Db	2560		CCCACTCTACACTGTTTCCACAGGCGCCAGCTGCGGGGGAGACCGTGGCCCGGGCC	2619
QY	3016		CTCTTCTGGAGTCTCCCAAGGACTCTAGACACTTGGAGCTGTGGAGCACAGCTCTGTGG	3075
Db	2620		CTCTTCTGGAGTCTCCCAAGGACTCTAGACACTTGGAGCTGTGGAGCACAGCTCTGTGG	2679
QY	3076		GGGGAGGCGCTCTCATCAACCCAGTGTCTCAGGCGGGGAGGCGGAGTGAAGTGAAGTGA	3135
Db	2680		GGGGAGGCGCTCTCATCAACCCAGTGTCTCAGGCGGGGAGGCGGAGTGAAGTGAAGTGA	2739
QY	3136		TTTCCCTTGGGACATGTCAGACCTGACAGCGGTCCAAATAGAGCCCTTCCGACGCTC	3195
Db	2740		TTTCCCTTGGGACATGTCAGACCTGACAGCGGTCCAAATAGAGCCCTTCCGACGCTC	2799
QY	3196		CCACCCCACTGACGCTCCCGTGGAGCCATCCACAGCGAGGGGAGTGGGTGACG	3255
Db	2800		CCACCCCACTGACGCTCCCGTGGAGCCATCCACAGCGAGGGGAGTGGGTGACG	2859
QY	3256		CTGCGGCGCCCTTGGACCAACATCAACCTTCCAGCTTGGGGTGGGTATCATATCCCCCTG	3315
Db	2860		CTGCGGCGCCCTTGGACCAACATCAACCTTCCAGCTTGGGGTGGGTATCATATCCCCCTG	2919
QY	3316		CAGGGCTTGGCTTCCACAGGTCGCCAGGAGCCATGCGCCCTGCTGGCTGGGCG	3375
Db	2920		CAGGGCTTGGCTTCCACAGGTCGCCAGGAGCCATGCGCCCTGCTGGCTGGGCG	2979
QY	3376		CTAAACAAAGGTTGAGAGGCGGAGGAGCTGTCTTGGGACGATGGAGAGCTTGGAA	3435
Db	2980		CTAAACAAAGGTTGAGAGGCGGAGGAGCTGTCTTGGGACGATGGAGAGCTTGGAA	3039
QY	3436		GTGCTGGAGCGAGGGCGCTTACACAGGTCTCTTCTTGGGCGAGGATTAACAGATCGTG	3495
Db	3040		GTGCTGGAGCGAGGGCGCTTACACAGGTCTCTTCTTGGGCGAGGATTAACAGATCGTG	3099
QY	3496		ATGAGCTGTGATGAGTGTGACCAAGTGTGGGAGCTGGCTGACGCTGCAAGGATGACTGTC	3555
Db	3100		ATGAGCTGTGATGAGTGTGACCAAGTGTGGGAGCTGGCTGACGCTGCAAGGATGACTGTC	3159

QY 3556 CTGGGCGTGGCCACCGCGCCCGCCAGCAGGTCTCTCTCAACGGGTGTCCTGTCTCCAACTTC 3615
DB 3160 CTGGGCGTGGCCACCGCGCCCGCCAGCAGGTCTCTCTCAACGGGTGTCCTGTCTCCAACTTC 3219
QY 3616 ACTACAGCCCGACACACCAAGTCTCTGGACATCTGTCTCGCTGTGTGATGGAGAGCAG 3675
DB 3220 ACTACAGCCCGACACACCAAGTCTCTGGACATCTGTCTCGCTGTGTGATGGAGAGCAG 3279
QY 3676 TTCTCGTCAGCTGGTGT 3694
DB 3280 TTCTCGTCAGCTGGTGT 3298

RESULT 2

US-09-949-016-3656
; Sequence 3656, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3656
; LENGTH: 3847
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-3656

Query Match 71.4%; Score 2644.6; DB 4; Length 3847;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2650; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1036 GAAACCGGTGACACCCCGCGCTCCAGAGCAGTCCACACAGTGCAGCGTCCCGCC 1095
DB 639 GATGCCAGGCACACCCCGCGCTCCAGAGCAGTCCACACAGTGCAGCGTCCCGCC 698
QY 1096 AACAGCGGTTCGATTGCGCCCTGACAAAGGCCATCACCCAGAAACAGTGCAGGCCGC 1155
DB 699 AACAGCGGTTCGATTGCGCCCTGACAAAGGCCATCACCCAGAAACAGTGCAGGCCGC 758
QY 1156 GGCTGCTGCTACATCCCTGCAAGCAGGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGG 1215
DB 759 GGCTGCTGCTACATCCCTGCAAGCAGGGGCTGCAGGGAGCCAGATGGGGCAGCCCTGG 818
QY 1216 TGCTTCTCCACCCAGCTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAAATG 1275
DB 819 TGCTTCTCCACCCAGCTACCCAGCTACAGCTGAGAACTGAGCTCTCTGAAATG 878
QY 1276 GGCTACAGCGCCACCTGACCCGTACCCACCCACCTTCTCCCAAGGACATCTCTGACC 1335
DB 879 GGCTACAGCGCCACCTGACCCGTACCCACCCACCTTCTCCCAAGGACATCTCTGACC 938
QY 1336 CTGGCGTGGACGTGATGATGAGACTGAGAACCCGCTCCACTTACGATCAAGATCCA 1395
DB 939 CTGGCGTGGACGTGATGATGAGACTGAGAACCCGCTCCACTTACGATCAAGATCCA 998
QY 1396 GCTAACAGCGCTTACGAGGTGCCCTTGGAGACCCCGCTGTGCACAGCCGGCAGCGTCC 1455
DB 999 GCTAACAGCGCTTACGAGGTGCCCTTGGAGACCCCGCTGTGCACAGCCGGCAGCGTCC 1058
QY 1456 CCACCTTACAGCGTGGAGTCTCCGAGGAGCCCTTCGGGGTGTATCGTGACCCGGCAGCTG 1515

DB 1059 CCACCTTACAGCGTGGAGTTCCTCGAGGAGCCCTTCGGGGTGTATCGTGACCCGGCAGCTG 1118
QY 1516 GACGCGCGGTGCTGCTGAACACGACGCTGGCGCCCTGTCTTTTTCGGAACGAGTTCCTT 1575
DB 1119 GACGCGCGGTGCTGCTGAACACGACGCTGGCGCCCTGTCTTTTTCGGAACGAGTTCCTT 1178
QY 1576 GAGCTGTCCACCTCGCTGCTGAGTATATACAGGCTCGCCGAGCAGCTCAGTCC 1635
DB 1179 GAGCTGTCCACCTCGCTGCTGAGTATATACAGGCTCGCCGAGCAGCTCAGTCC 1238
QY 1636 CTGATGCTCAGCACCGCTGGACAGGATCACCTGTGGAACCGGACCTTGGCCCAAG 1695
DB 1239 CTGATGCTCAGCACCGCTGGACAGGATCACCTGTGGAACCGGACCTTGGCCCAAG 1298
QY 1696 CCCGCTGCGAACCTCTACCGGTCTCACCTTTTCTACTGCGCTGAGAACCGCGGTG 1755
DB 1299 CCCGCTGCGAACCTCTACCGGTCTCACCTTTTCTACTGCGCTGAGAACCGCGGTG 1358
QY 1756 GCACACGGGTGTTCTGCTAAACAGCAATGCATGAGTGTGCTCTGAGCAGCGCCCT 1815
DB 1359 GCACACGGGTGTTCTGCTAAACAGCAATGCATGAGTGTGCTCTGAGCAGCGCCCT 1418
QY 1816 GCCCTTAGCTGAGGTGACAGGTGCGATCCTGGATGCTACATCTTCTGGGCCAGAG 1875
DB 1419 GCCCTTAGCTGAGGTGACAGGTGCGATCCTGGATGCTACATCTTCTGGGCCAGAG 1478
QY 1876 CCCAAGAGCGTGGTGCAGCAGTACCTGACGTGTTGGGATACCCGTTCTATGCCCATAC 1935
DB 1479 CCCAAGAGCGTGGTGCAGCAGTACCTGACGTGTTGGGATACCCGTTCTATGCCCATAC 1538
QY 1936 TGGGCGCTGGGTTCACCTGTGCGCTGGGCTACTCTCCACCGCTATACCCGCCAG 1995
DB 1539 TGGGCGCTGGGTTCACCTGTGCGCTGGGCTACTCTCCACCGCTATACCCGCCAG 1598
QY 1996 GTGGTGGAGAACATGACAGGCGCCACTTCCCTGGAGCTCCAAATGGAACGACCTGGAC 2055
DB 1599 GTGGTGGAGAACATGACAGGCGCCACTTCCCTGGAGCTCCAAATGGAACGACCTGGAC 1658
QY 2056 TACATGGACTCCCGAGGAGCTTTCAGTTTCAACAAAGGATGGCTTCGCGGACTTCCCGGCC 2115
DB 1659 TACATGGACTCCCGAGGAGCTTTCAGTTTCAACAAAGGATGGCTTCGCGGACTTCCCGGCC 1718
QY 2116 ATGTTGAGGAGTGCACAGGCGCGCGCTACATGATGATGATGATGATGATGATGATGATG 2175
DB 1719 ATGTTGAGGAGTGCACAGGCGCGCGCTACATGATGATGATGATGATGATGATGATGATG 1778
QY 2176 AGCAGCTGGGCGCTTCGCGGAGCTACAGGCGCTACGAGGCTTCGCGGAGGCGGT 2235
DB 1779 AGCAGCTGGGCGCTTCGCGGAGCTACAGGCGCTTCGCGGAGGCGGTTCGCGGAGGCGGT 1838
QY 2236 TTTATCACCAACGAGACCGCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCC 2295
DB 1839 TTTATCACCAACGAGACCGCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCC 1898
QY 2296 TTTCCCGGACTTCAACAAACCCACAGCCCTTGGGCTGTTGGGAGGACATGTTGCTGATTC 2355
DB 1899 TTTCCCGGACTTCAACAAACCCACAGCCCTTGGGCTGTTGGGAGGACATGTTGCTGATTC 1958
QY 2356 CATGACGAGTGCCTTGAACGCGCTTGGATGATGATGATGATGATGATGATGATGATGATG 2415
DB 1959 CATGACGAGTGCCTTGAACGCGCTTGGATGATGATGATGATGATGATGATGATGATGATG 2018
QY 2416 AGAGGCTCTGAGGACGGCTGCGCCCAACAAATGAGCTGGAGAACCCACTTACGTGCTGGG 2475
DB 2019 AGAGGCTCTGAGGACGGCTGCGCCCAACAAATGAGCTGGAGAACCCACTTACGTGCTGGG 2078
QY 2476 GTGGTGGGGGACCTTCCAGCGCGCCACCATCTGTGCTTCCAGCCACAGTTCCTTCTCC 2535
DB 2079 GTGGTGGGGGACCTTCCAGCGCGCCACCATCTGTGCTTCCAGCCACAGTTCCTTCTCC 2138
QY 2536 ACACACTAACCTGACAAACCTCTACGCGCTGACCGAGGACCATGCTCTCCACAGGGCG 2595
DB 2139 ACACACTAACCTGACAAACCTCTACGCGCTGACCGAGGACCATGCTCTCCACAGGGCG 2198

```
QY 2596 CTGGTGAAGGCTCGGGGACACGCCATTGTGATCTCCGCTCGACCTTTGCTGGCCAC 2655
Db 2199 CTGGTGAAGGCTCGGGGACACGCCATTGTGATCTCCGCTCGACCTTTGCTGGCCAC 2258
QY 2656 GCGCATACCGCGGCCACTGGACGGGGACGTGTGGAGTCTCTGGAGCAGTCCGCTCC 2715
Db 2259 GCGCATACCGCGGCCACTGGACGGGGACGTGTGGAGTCTCTGGAGCAGTCCGCTCC 2318
QY 2716 TCGTGCCAGAAATCTGCAGTTTAACTCTGCGGGGTGCTCTGGTGGGGCCACGTC 2775
Db 2319 TCGTGCCAGAAATCTGCAGTTTAACTCTGCGGGGTGCTCTGGTGGGGCCACGTC 2378
QY 2776 TCGGGCTTCTGGGAAACCTCAGAGGAGTGTGTGCGCTGGACCCAGCTGGGGCC 2835
Db 2379 TCGGGCTTCTGGGAAACCTCAGAGGAGTGTGTGCGCTGGACCCAGCTGGGGCC 2438
QY 2836 TTCTACCCCTTATGCGGAACCAACAGCCTGTCTAGTCTGCCAGAGCGGTACAGC 2895
Db 2439 TTCTACCCCTTATGCGGAACCAACAGCCTGTCTAGTCTGCCAGAGCGGTACAGC 2498
QY 2896 TTTCAGGAGCGCGCCAGCAGGCCATGAGGAAGGCCCTCACCTCGCTACGACTCCTC 2955
Db 2499 TTTCAGGAGCGCGCCAGCAGGCCATGAGGAAGGCCCTCACCTCGCTACGACTCCTC 2558
QY 2956 CCCCACCTCTACACACTGTTCACACGAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 3015
Db 2559 CCCCACCTCTACACACTGTTCACACGAGGCCACGTCGCGGGGAGACCGTGGCCCGGCC 2618
QY 3016 CTCCTCTGGAGTTCGCCAAGACTTAGCAGCTGGAGCTGTGGACCAAGCTCTGTGG 3075
Db 2619 CTCCTCTGGAGTTCGCCAAGACTTAGCAGCTGTGGAGCTGTGGACCAAGCTCTGTGG 2678
QY 3076 GGGAGGCCCTCTCATATACCCAGTGTCTCAGGCGGGAGGAGGCGGAGTGAAGTACTGCTAC 3135
Db 2679 GGGAGGCCCTCTCATATACCCAGTGTCTCAGGCGGGAGGAGGCGGAGTGAAGTACTGCTAC 2738
QY 3136 TTCCCTCTGGGACATGTGACAGCTGTGACAGCGGTCCAAATAGAGCCCTTGGCAGCCTC 3195
Db 2739 TTCCCTCTGGGACATGTGACAGCTGTGACAGCGGTCCAAATAGAGCCCTTGGCAGCCTC 2798
QY 3196 CCACCCCACTGTGAGCTCCCGTGTAGCCAGGCATCCACAGCAGGGGAGTGGGTGAGC 3255
Db 2799 CCACCCCACTGTGAGCTCCCGTGTAGCCAGGCATCCACAGCAGGGGAGTGGGTGAGC 2858
QY 3256 CTGGCGGCCCTCTGACACCATCAAGTCCACCTCCGCGGTGGGTACATCATCCCCCTG 3315
Db 2859 CTGGCGGCCCTCTGACACCATCAAGTCCACCTCCGCGGTGGGTACATCATCCCCCTG 2918
QY 3316 CAGGGCCCTTGGCTTCAACACAGAGTCCCGCCAGCAGCCCATGGCCCTGTGGCTGTGGCC 3375
Db 2919 CAGGGCCCTTGGCTTCAACACAGAGTCCCGCCAGCAGCCCATGGCCCTGTGGCTGTGGCC 2978
QY 3376 CTAAACAGAGTGGAGAGCGCCGAGGGAGCTGTCTGGGACATGGAGAGCCTGGAA 3435
Db 2979 CTAAACAGAGTGGAGAGCGCCGAGGGAGCTGTCTGGGACATGGAGAGCCTGGAA 3038
QY 3436 GTGCTGGAGCGAGGGCCCTACACAGAGTCACTCTCTGGCCAGGAATAACACGATCGTG 3495
Db 3039 GTGCTGGAGCGAGGGCCCTACACAGAGTCACTCTCTGGCCAGGNATTAACAGATCGTG 3098
QY 3496 AATGAGCTGGTATGATGACAGTGGAGGAGTGGCTCTCCAGCGGTGCTCTCCAACTTC 3555
Db 3099 AATGAGCTGGTATGATGACAGTGGAGGAGTGGCTCTCCAGCGGTGCTCTCCAACTTC 3158
QY 3556 CTGGGGGTGGCCACCGCGGCCAGCAGGTCTCTCCAGCGGTGCTCTCCAACTTC 3615
Db 3159 CTGGGGGTGGCCACCGCGGCCAGCAGGTCTCTCCAGCGGTGCTCTCCAACTTC 3218
QY 3616 ACTACAGCCCGACACCAAGGTCCTGGAGCATCTGTGCTCTGCTGTGATGGAGAGCAG 3675
Db 3219 ACTACAGCCCGACACCAAGGTCCTGGAGCATCTGTGCTCTGCTGTGATGGAGAGCAG 3278
```

```
QY 3676 TTTCTCGTCAGCTGGTGT 3694
Db 3279 TTTCTCGTCAGCTGGTGT 3297

RESULT 3
US-09-397-787-175
; Sequence 175, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.466C2
; CURRENT APPLICATION NUMBER: US/09/397,787
; CURRENT FILING DATE: 1999-09-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 175
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-397-787-175

Query Match 10.0%; Score 370.6; DB 3; Length 449;
Best Local Similarity 98.9%; Pred No. 3.4e-67;
Matches 373; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 63 CGGATCTTACTCGCGGGAGAGAACCCAGCCCGTCCCGGAAACCGAGTCCGGAGA 122
Db 73 CGGCAAGTACTCGCGGGAGAGAACCCAGCCCGTCCCGGAAACCGAGTCCGGAGA 132
QY 123 GGAGTTCCGCATGGAGAGTTGAACCCAGCTGTGGAGAGAGCCCGAGCTGCATCTTCC 182
Db 133 GGAGTTCCGCATGGAGAGTTGAACCCAGCTGTGGAGAGAGCCCGAGCTGCATCTTCC 192
QY 183 TCCCGTAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 242
Db 193 TCCCGTAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 252
QY 243 CTGGAAGAACTAAAGCTTGACGGCTTGGACGAAGATGGGAGAGAGGAGGAGCACTCAT 302
Db 253 CTGGAAGAACTAAAGCTTGACGGCTTGGACGAAGATGGGAGAGAGGAGGAGCACTCAT 312
QY 303 AGCGAACCTCAATGTCTCATCTTGGCCCAAGTATGCTTGGAGAGAGAGGAGGAGCGCTCGCA 362
Db 313 AGCGAACCTCAATGTCTCATCTTGGCCCAAGTATGCTTGGAGAGAGAGGAGGAGCGCTCGCA 372
QY 363 GGTGACCACTCACTCCCTCAGTGGCACCCAGGAGACCGGCTGGATGACCCCGAGGCTGA 422
Db 373 GGTGACCACTCACTCCCTCAGTGGCACCCAGGAGACCGGCTGGATGACCCCGAGGCTGA 432
QY 423 AAAGCTGTGGCACAAGG 439
Db 433 AAAGCTGTGGCACAAGG 449
```

```
RESULT 4
US-08-430-925A-3
; Sequence 3, Application US/08430925A
; Patent No. 5763252
; GENERAL INFORMATION:
; APPLICANT: Skadsen, Ronald W
; APPLICANT: Tibbot, Brian K
; TITLE OF INVENTION: Cloned Alpha-Glucosidase from Barley
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
```

STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,925A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296.93171
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2752 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 42..2675
US-08-430-925A-3

Query Match 10.0%; Score 369.8; DB 1; Length 2752;
Best Local Similarity 53.6%; Pred. No. 8.1e-67;
Matches 966; Conservative 0; Mismatches 772; Indels 63; Gaps 7;
QY 1687 GCGCCACGCCCGGTGCGAACCTCTACGGGTCTACCCCTTCTACCTGGGGTGGAGGAC 1746
DB 648 GCGTCTAGTGGAGCTCAACCTCTACGGGTGCGACCCCTTCTACATGGAGCGTGGGGCA 707
QY 1747 GCGGGTGGGACACGCGGGTGTCTCGCTAAACAGCAATGCCATGATGTGGTCTGCGAG 1806
DB 708 CCGGGAGCCGCGACGCGGTGCTCTCTGCTCAGTAGCAAGCATGAGCGTG---CTCTAC 764
QY 1807 CCGAGCCCTGCCCTTAGCTGGAGGTGCGACAGGTGGGATCTGGAGTCTACATCTTCCTG 1866
DB 765 GCGGGTCTCTACGTCACTTACAGGTCTATCGGGGGGTCTCTGACTTCTACTTCTTCGCC 824
QY 1867 GCGCCAGAGCCCAAGAGCGTGTGCGAGAGTACCTGGAGTGTGGGATACCGGTTCATG 1926
DB 825 GCGCCCAACCCCTCGCCGTGCTGCGACAGTACACCCAGTCTATCGCCCGCCCTGCCCCA 884
QY 1927 CCGCCATCTGGGGCTGGGCTTCCACCTGTGCGGTGGGGTCTCTCCACCGCTATC 1986
DB 885 ATGCGGTACTGTGCTTCTGGGTTCACACAGTGGCGGTACGGGTACTGAACTGTCTGAC 944
QY 1987 ACCCGCAGGTGTGGAGAACATGACAGGGGCCCACTTCCCTCGACAGCTCCAAATGGAAC 2046
DB 945 CTGGAGCGTGTGTGGCCCGATACGCCAAGGCCCGGATCCCGTGGAGGTGATGAGACC 1004
QY 2047 GACCTGGACTACATGACTCCCGGAGGACTTCACTTCAACAGAGATGGCTT-----C 2100
DB 1005 GATATCGACTACATGACCGGGTTCAGAGACTTCACCTTGGACCGTGTCACTTCAACCGCC 1064
QY 2101 CCGGACTTCCCGCATGTGCGAGGAGTGCACAGGGCGCGCGCTACATGATGATC 2160
DB 1065 GCGAGCTCCGCGCGTGTGCTGACCGCGTTCACCGAAGCCCGAAGATACGCTCTCATC 1124
QY 2161 GTGGATCTCGCATCAGCAGCTCGGGCCCTTCCCGGAGGTACAGGCCCTTACAGCAGGGT 2220
DB 1125 CTAGACCCAGGATCCGGGTGGACCCATCGAGCGACGTACGGGACGTTGTCGCCGG 1184
QY 2221 CTGCGAGGGGGTTCATATCAACAGAGACCGGCCACCGCTGATGTGGAGGTATGG 2280

DB 1185 ATGACAGGACATCTTCTCT---GAAGCGGAAGCGCAAAACTTCTGTCGCAACGTGTGG 1241
QY 2281 CCGGGTCCACTGCTTCCCGACTTCAACACCCACAGCCTTGGCTTGGTGGAGGAC 2340
DB 1242 CCGGGCGAGCTTACTTCTCCCGACTTATGACCCAGCGCGCGAGTCTTGGGCGCG 1301
QY 2341 ATGGTGGTGAATTCATGACAGGTGCGCTTCGAGCGGTGTGGATTGACATGACGAG 2400
DB 1302 GAGATCTCTCTCTTCCCGCGGACCAATCCCGGTGCGAGCGGTGTGGATGACATGAACGAG 1361
QY 2401 CCTTCAACTTCAATCAGAGGCTCTGAGGACGCTGCCCAACAAATGAGCTGAGAGAACCA 2460
DB 1362 ATCTCAACTTCTACA---ACCGGAGCCATGAAGCGCTGACGACCGCCGTTACCGG 1418
QY 2461 CCTACGTGCTGGGGTGTGGGGGACCTTCCAGGGCGGCACCAATCTGTGTCTCCAGC 2520
DB 1419 ATCAACACGACGCGGACGGCGCGCCCATCAACAAACAGACGCTGCGCGCTGCGCGTG 1478
QY 2521 GACCAAGTTCCTTCCACACACTCAACCTGCAACCTTACGGCTTACCGGCTGACGAGCCATC 2580
DB 1479 CACTACGGGGGGTCAACCGAGTACGAGGAGCAACCTTCTCGGCTTCTCGAGGGCGCGC 1538
QY 2581 GCTTCCACAGGGCGCTGCTGAAGGCTCGGGGACACGCGCCATTTGTGATCTCCCGCTCG 2640
DB 1539 GCAACGGGGCGGGGTGCTGAGGAGACCCGCGCGCGCTTCTCGGCTTCTCGAGCAGGTCC 1598
QY 2641 ACCTTTGTGCGCCACGCGCGATACGCGGCCACTGGAACGGGGACGTGTGGAGCTCTCTGG 2700
DB 1599 ACCTTGTGGGCTCGGGCGCTACACCGCTTACTGGACTGGCGCAACGCGCAACGTTGG 1658
QY 2701 GAGCAGCTCGCTCTCGTGCAGAAATCTTGCAGTTTAACTGTGGGGGTGCTCTCTG 2760
DB 1659 GCGCAGCTCGCTTCTTCAACACCATGCTCAGCTTCGGCTTCTTCGGCATGGCCATG 1718
QY 2761 GTCGGGGCGAGCTCTGCGGCTTCTTGGGCAACACTCAGAGAGCTGTGTGCGCTGG 2820
DB 1719 ATCGGGCGCGCATCTGCGGTTCAACGCAACACGACAGAGAGCTTCTCGGTCTGGTGG 1778
QY 2821 ACCCAGCTGGGGCGCTTCTACCCCTTTCATGCGGAACCAACAGCCTGTCTAGTCTGGCC 2880
DB 1779 ATCCAGCTCGAGCTTCTTACCCCTTTCGAGGGACCACTTCGCGCATCTTACCGTCCGG 1838
QY 2881 CAGGAGCGGTACAGCTTACGAGCGCGGCCAGAGGCCATGAGGAGGCCCTTACCCCTG 2940
DB 1839 CGAGAGTTGTA---CCTGTGGCGCTCGGTGGCGGCTCGGGCAGGAAGCGCTCGGGCTC 1895
QY 2941 CGCTACGCACTCTCCCGCACCTTACACACTTTCACAGCGCTCATGTACAGGGCGCATGACGGGGCG 3000
DB 1896 CGGTACAGCTCTCTCTTACTTCTACAGCTCATGTACAGGGCGCATGACGGGGCG 1955
QY 3001 ACCGTGGCGCGCGCTCTTCTCTGGAGTTCCCAAGGACTCTAGCACCTGGACTGGAC 3060
DB 1956 CCAATCGCGCGCGCTTCTTCTTCTTACCCGACGAGCTCGCCACGTTACGGCGTGGAC 2015
QY 3061 CACCAAGCTCTGTGGGGAGGCGCTTCTATCAACCCAGTGTCTCAGGCGCGGGAAGGCC 3120
DB 2016 AGACAGTCTCTCTCGCGCGCGGGTCTCTGTTTCTGCGGTGTCTGAGCGCGGCGCGACA 2075
QY 3121 GAAGTGACTGGTACTTCTTCTTGGGACATGCTAGCAGCTTCAGAGCGGTGCAATAGAG 3180
DB 2076 ACCGTGAGCGCTACTTCTTCTCGCGCGCGGTGGTA-----CAGACTTACGACTACTC 2128
QY 3181 GCGCTTGGCAGCTCTCCACCCCACTGTCAGCTCTCCCGTGGAGCGACCATCCACAGCGAG 3240
DB 2129 CTGCGCGTGCACGCGGACCGGCAAGCACTC----- 2162
QY 3241 GGGCAGTGGGTGACGCTGCGCGCGCGCTTGGACACCATCAAGCTCACCTCGGGGTGGG 3300
DB 2163 -----AGGCTGCGCGCGCGCGCGGACAGGTGAAGTGTGACCTGACCGCGCGC 2210
QY 3301 TACATCATCCCCCTGACGGGCCCTGGCTTCAACACAGAGTCCCGCGCAGCGCCATG 3360
DB 2211 ACCATCTCTCCCGTGCAGCAGCGCTGACTACTGTCGCGCGCGCGCGGACCGCGTTC 2270


```
QY 2333 GGGAGGACATGGTGGCTGAGTTCCATGACCAAGTGCCTTCGACGGCTTGTGGATTGACA 2392
Db 61 GGGCAAAATGAATGCAGTATTTTCCATCAAGAAGTCAATATATGATGACTTTTGGATTGACA 120
QY 2393 TGAAGCAGCTTCCAACTTCATCAGAGGCTCTGAGGAGCGCTGCCCAACAATGAGCTGG 2452
Db 121 TGAATGAAGTTTCCAGCTTATTTCAAGGTTCAAAAAGGATGTAATGTAACAATGGA 180
QY 2453 AGAACCCACCTACGTGCTGGGGTGGTGGGGGACCTCCAGGCGGCCACCATCTGTG 2512
Db 181 ATTATCCACGGTTTACTCTGTATTTCTTGACAAACTCATGTATTTCCAAAACAATTTGCA 240
QY 2513 CTTCCAGCACCAGATTTTCTTCCACACACTCAAACTGCAACACTCTACGGCTGACGG 2572
Db 241 TGGATCTGTGCAAGACTGGGGTAAACAGTATGTTTCATAGCTCTATGANTACGCA 300
QY 2573 AAGCCATCGCTCCCAACAGGGCGCTGGTGAAGS---CTCGGGGGACACGCCCATTTGGA 2629
Db 301 TGGCTATAGCCACAGAGCAAGCTGTACAAAAGTTTTTCTTAATAGAAAGCTTCATTC 360
QY 2630 TCTCCCGCTCGACCTTTGTGTCGCCACCGCGGATACGCCGCCACTGGACGGGGGACGTGT 2689
Db 361 TTACCGCTCAACATTTGTGATCTGGAAGACATGCTCTCATTTGGTTAGGAGACAATA 420
QY 2690 GGAGCTCTGGGAGCAGCTGCGCTCTCCGTGCGCAAAATCTCGCAGATTTAACTGCTGG 2749
Db 421 CTGCTTTCATGGGAACAAATGGAATGCTCTATAACTGGAATGCTGGAGTTTCAGTTGTTG 480
QY 2750 GGGTGCTCTGTCGGGGCGGACGCTGCGGCTTCTCTGGGCAACACTCAGAGGAGCTGT 2809
Db 481 GAATACCTTTGGTGAGCAGACATCTGTGAAATTTGTGCTGAATTTGAGCAACAGAAACTTT 540
QY 2810 GTGTGCGCTGGACCCAGCTGGGGGCTCTTACCCCTTCTATGCGGAACCAACAGCCTGC 2869
Db 541 GCAGAGATGGATGCAACTTGGGGCATTTTATCCATTTTCCAGAAACCATATTTCTGAGC 600
QY 2870 TCAGTCTGCCAGAGCGGTACAGCTTCAGGAGCGGCGCCAGCAGG-----CAATGA 2923
Db 601 GATATGAACATCAGGATCTGCAATTTTGGGCGAATTCACATTTTGGTTAAATCATCAA 660
QY 2924 GGAAGCCCTCACCTCGGCTAGCACTCTCCGCCACTCTACACACTGTTCACCCAGG 2983
Db 661 GGCAATTTAACTATTCGCTACACTTATACCTTATACCTTCTACACTCTGTTTATAAG 720
QY 2984 CCCACGTGCGGGGGAGACGCTGGCGCGGCTCTTCTCTGGAGTTTCCCAAGGACTCTA 3043
Db 721 CCATGTGTTTGAGAGAAACAGTAGCAAGACCAAGTCTCTCATGAGTTTATGAGGATACGA 780
QY 3044 GCACCTGAGCTGGACCAACAGCTCTCTGTGGGGAGGCGCTCTCATCACCACGAGTGC 3103
Db 781 ACAGCTGGATTGAGGACACTGAGTTTGTGGGGCGCTGCAATTTACTTACTCTCTGTTTC 840
QY 3104 TCCAGCGCGGAGCGGAGGCTGACCTGCTACTCTCCCTTGGGCACATGTTAGCA 3158
Db 841 TAAACAGGAGCAGATATCTGTAGTGCCTATCCCTCATCTGCTATTTGGTATGA 895

RESULT 9
US-09-949-016-18094
; Sequence 18094, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18094
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-18094

Query Match 4.7%; Score 175.2; DB 4; Length 601;
Best Local Similarity 77.8%; Pred. No. 7.6e-27;
Matches 210; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

QY 3306 CATCCCTTCGAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 3365
Db 217 CTTCCCTTTCCAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 276
QY 3366 GCGTGTGCGCTTAAACCAAGGTGGAGAGCGCGGAGAGCTGTTCTGGGACATGGAGA 3425
Db 277 GCGTGTGCGCTTAAACCAAGGTGGAGAGCGCGGAGAGCTGTTCTGGGACATGGAGA 336
QY 3426 GAGCCTGGAAGTGTGGAGCGAGGGGCTTACACACAGAGTCTTCTGCGCCAGGAATAA 3485
Db 337 GAGCCTGGAAGTGTGGAGCGAGGGGCTTACACACAGAGTCTTCTGCGCCAGGAATAA 396
QY 3486 CACGATCTGTAATGAGTGTGTACGTGTACCACTGAGGAGCTGGCTTGCAGCTGCAGAA 3545
Db 397 GAGTCTCGGGGCTGCTCAGGCTGTGTGGGCGGCGCTCGGGGTGTGAGAAAGGGGTGAG 456
QY 3546 GTGAGTCTCTGGGCGTGGCCACCGCGCC 3575
Db 457 GGGACCTGGGCTTGGGGGTCCACCATGGC 486

RESULT 10
US-09-949-016-128902
; Sequence 128902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128902
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-128902

Query Match 4.7%; Score 175.2; DB 4; Length 601;
Best Local Similarity 77.8%; Pred. No. 7.6e-27;
Matches 210; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

QY 3306 CATCCCTTCGAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 3365
Db 217 CTTCCCTTTCCAGGCGCTTGGCTTCAACACACAGAGTCCCGCCAGCAGCCCATGGCCCT 276
QY 3366 GCGTGTGCGCTTAAACCAAGGTGGAGAGCGCGGAGAGCTGTTCTGGGACATGGAGA 3425
Db 277 GCGTGTGCGCTTAAACCAAGGTGGAGAGCGCGGAGAGCTGTTCTGGGACATGGAGA 336
QY 3426 GAGCCTGGAAGTGTGGAGCGAGGGGCTTACACACAGAGTCTTCTGCGCCAGGAATAA 3485
```

```
QY 2333 GGGAGGACATGGTGGCTGAGTTCCATGACCAAGTGCCTTCGACGGCTTGTGGATTGACA 2392
Db 61 GGGCAAAATGAATGCAGTATTTTCCATCAAGAAGTCAATATATGATGACTTTTGGATTGACA 120
QY 2393 TGAAGCAGCTTCCAACTTCATCAGAGGCTCTGAGGAGCGCTGCCCAACAATGAGCTGG 2452
Db 121 TGAATGAAGTTTCCAGCTTATTTCAAGGTTCAAAAAGGATGTAATGTAACAATGGA 180
QY 2453 AGAACCCACCTACGTGCTGGGGTGGTGGGGGACCTCCAGGCGGCCACCATCTGTG 2512
Db 181 ATTATCCACGGTTTACTCTGTATTTCTTGACAAACTCATGTATTTCCAAAACAATTTGCA 240
QY 2513 CTTCCAGCACCAGATTTTCTTCCACACACTCAAACTGCAACACTCTACGGCTGACGG 2572
Db 241 TGGATCTGTGCAAGACTGGGGTAAACAGTATGTTTCATAGCTCTATGANTACGCA 300
QY 2573 AAGCCATCGCTCCCAACAGGGCGCTGGTGAAGS---CTCGGGGGACACGCCCATTTGGA 2629
Db 301 TGGCTATAGCCACAGAGCAAGCTGTACAAAAGTTTTTCTTAATAGAAAGCTTCATTC 360
QY 2630 TCTCCCGCTCGACCTTTGTGTCGCCACCGCGGATACGCCGCCACTGGACGGGGGACGTGT 2689
Db 361 TTACCGCTCAACATTTGTGATCTGGAAGACATGCTCTCATTTGGTTAGGAGACAATA 420
QY 2690 GGAGCTCTGGGAGCAGCTGCGCTCTCCGTGCGCAAAATCTCGCAGATTTAACTGCTGG 2749
Db 421 CTGCTTTCATGGGAACAAATGGAATGCTCTATAACTGGAATGCTGGAGTTTCAGTTGTTG 480
QY 2750 GGGTGCTCTGTCGGGGCGGACGCTGCGGCTTCTCTGGGCAACACTCAGAGGAGCTGT 2809
Db 481 GAATACCTTTGGTGAGCAGACATCTGTGAAATTTGTGCTGAATTTGAGCAACAGAAACTTT 540
QY 2810 GTGTGCGCTGGACCCAGCTGGGGGCTCTTACCCCTTCTATGCGGAACCAACAGCCTGC 2869
Db 541 GCAGAGATGGATGCAACTTGGGGCATTTTATCCATTTTCCAGAAACCATATTTCTGAGC 600
QY 2870 TCAGTCTGCCAGAGCGGTACAGCTTCAGGAGCGGCGCCAGCAGG-----CAATGA 2923
Db 601 GATATGAACATCAGGATCTGCAATTTTGGGCGAATTCACATTTTGGTTAAATCATCAA 660
QY 2924 GGAAGCCCTCACCTCGGCTAGCACTCTCCGCCACTCTACACACTGTTCACCCAGG 2983
Db 661 GGCAATTTAACTATTCGCTACACTTATACCTTATACCTTCTACACTCTGTTTATAAG 720
QY 2984 CCCACGTGCGGGGGAGACGCTGGCGCGGCTCTTCTCTGGAGTTTCCCAAGGACTCTA 3043
Db 721 CCATGTGTTTGAGAGAAACAGTAGCAAGACCAAGTCTCTCATGAGTTTATGAGGATACGA 780
QY 3044 GCACCTGAGCTGGACCAACAGCTCTCTGTGGGGAGGCGCTCTCATCACCACGAGTGC 3103
Db 781 ACAGCTGGATTGAGGACACTGAGTTTGTGGGGCGCTGCAATTTACTTACTCTCTGTTTC 840
QY 3104 TCCAGCGCGGAGCGGAGGCTGACCTGCTACTCTCCCTTGGGCACATGTTAGCA 3158
Db 841 TAAACAGGAGCAGATATCTGTAGTGCCTATCCCTCATCTGCTATTTGGTATGA 895

RESULT 9
US-09-949-016-18094
; Sequence 18094, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
```


Db 380 GTGCAGAGCTGCACACAGGCGCGCGCTACATGATGATCGTGTGTGTCGCCCCACCA 439
QY 2179 AGCTCGGGCCCTG 2191
Db 440 CTGTGGGTCTTTG 452

RESULT 14
US-09-949-016-128889
; Sequence 128889, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128889
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-128889

Query Match 4.3%; Score 157.4; DB 4; Length 601;
Best Local Similarity 71.3%; Pred. No. 3.6e-23;
Matches 266; Conservative 1; Mismatches 17; Indels 89; Gaps 1;

QY 1908 TGTGGGATACCCGTTTCATCCCGCATCTGCGGCTTGGGCTTCCACCTGTGCGGTGGG 1967
Db 80 TGCAGGATACCCGTTTCATCCCGCATCTGCGGCTTGGGCTTCCACCTGTGCGGTGGG 139

QY 1968 CTACTCTCAGCGCTATCACCAGGAGTGTGGAGAACATGACAGGCGCCACTTCCC 2027
Db 140 CTACTCTCAGCGCTATCACCAGGAGTGTGGAGAACATGACAGGCGCCACTTCCC 199

QY 2028 CCT----- 2030
Db 200 CTGTGTAGTGGGTGGTGGCAGGGAGGCAAGGGCTGGCGGACGGCTCTCTCAG 259

QY 2031 -----GGAGTCCATGGAAACGACCTGGACTAC 2058
Db 260 GCCCCAGCAGACGGTCCCGTGTGGCTGCAGGACGTCARTGGAAACGACCTGGACTAC 319

QY 2059 ATGGACTCCGAGGAGCTTCAGTTCAACAGGATGGCTCCGGACTTCCGGCCATG 2118
Db 320 ATGGACTCCGAGGAGCTTCAGTTCAACAGGATGGCTCCGGACTTCCGGCCATG 379

QY 2119 GTGCAGAGCTGCACACAGGCGCGCGCTACATGATGATCGTGGATCTGTCCATCAGC 2178
Db 380 GTGCAGAGCTGCACACAGGCGCGCGCTACATGATGATCGTGGATCTGTCCATCAGC 439

QY 2179 AGCTCGGGCCCTG 2191
Db 440 CTGTGGGTCTTTG 452

RESULT 15
US-09-949-016-18077
; Sequence 18077, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18077
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-18077

Query Match 4.2%; Score 155; DB 4; Length 601;
Best Local Similarity 95.8%; Pred. No. 1.1e-22;
Matches 158; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1366 AACCGCTCCACTTCCAGATCAAGATCCAGTAAACAGCGCTACGAGTGCCCTTGGAG 1425
Db 287 AACCGCTCCACTTCTTSTAGATCAAGATCCAGTAAACAGCGCTACGAGTGCCCTTGGAG 346

QY 1426 ACCCGCGGTGTCCACAGCGCGGACCGTCCCACTTCTACAGCGTGGAGTTCTCCGAGGAG 1485
Db 347 ACCCGCGGTGTCCACAGCGCGGACCGTCCCACTTCTACAGCGTGGAGTTCTCCGAGGAG 406

QY 1486 CCCTTGGGGTGATCGTGCACCGGAGCTGGAGCGCGGCTGCTG 1530
Db 407 CCCTTGGGGTGATCGTGCACCGGAGCTGGAGCGCGGCTGCTG 451

Search completed: April 4, 2005, 16:02:28
Job time : 613 secs

	Qy	Qy	Db	Qy	Db	Qy
1	CTTACGCCCATGCGGGTCCGAGCGGGCTCTGTGGCTGCTCTGGCTCTCTCGCACCGTG	60	1	CTTACGCCCATGCGGGTCCGAGCGGGCTCTGTGGCTGCTCTGGCTCTCTCGCACCGTG	60	
61	CTCGGATCTTACTCTCGGGGAGAGAACACGAGCCGCTCCCGAAAACGCGAGTCCGGA	120	61	CTCGGATCTTACTCTCGGGGAGAGAACACGAGCCGCTCCCGAAAACGCGAGTCCGGA	120	
121	GAGGAGTTCGCGATGGAGAGATTGAACACGCTGTGGGAGAGAGGCCCAACGCACTGCATCTT	180				

121 GAGGAGTTCGCGATGAGAGGTTGAACACAGCTGTGGGAGAAAGCCCGACGACTGCTTT 180
181 CCTCCCGTGAGGCTGCGCGAGCTCCAGCTGATCTGAAGATACAGAGAGGAGCAACTC 240
181 CCTCCCGTGAGGCTGCGCGAGCTCCAGCTGATCTGAAGATACAGAGAGGAGCAACTC 240
241 GCCTGGAAGAACTAAAGCTTTGACGGCTTGACGAAAGATGGGAGAGGAAAGCGAGACTC 300
241 GCCTGGAAGAACTAAAGCTTTGACGGCTTGACGAAAGATGGGAGAGGAAAGCGAGACTC 300
301 ATACGCAACTCAATGTCTATTGTCCTTTGGCCAAAGTATGGTCTGGA CGGAAGAGAGCGCTCGG 360
301 ATACGCAACTCAATGTCTATTGTCCTTTGGCCAAAGTATGGTCTGGA CGGAAGAGAGCGCTCGG 360
361 CAGGTGACAGCAACTCCCTCAGTGCACCCAGGAAGAGGGCTGATGACCCAGGCTG 420
361 CAGGTGACAGCAACTCCCTCAGTGCACCCAGGAAGAGGGCTGATGACCCAGGCTG 420
421 GAAAAGCTGTGGCAACAGGCGAAGACTCTGGGAAATTTCTCCGGCGAAGAACTGGACAAG 480
421 GAAAAGCTGTGGCAACAGGCGAAGACTCTGGGAAATTTCTCCGGCGAAGAACTGGACAAG 480
481 CTCTGGCGGAGTTCCTGTGATCAACAAAGAAAGTTTCAAGGTACAACTCTCTGTGGAG 540
481 CTCTGGCGGAGTTCCTGTGATCAACAAAGAAAGTTTCAAGGTACAACTCTCTGTGGAG 540
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAGAGCTATTAGCCCTCGGACCTGAGCGAC 600
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAGAGCTATTAGCCCTCGGACCTGAGCGAC 600
601 ATCAAGGCGAGCTCTGCAACAGCAGGACACAGGAGCTGAAGAGAGAGCTGGCGAGCATC 660
601 ATCAAGGCGAGCTCTGCAACAGCAGGACACAGGAGCTGAAGAGAGAGCTGGCGAGCATC 660
661 AACGAGGCTGAGACCGCTGCGCAGGCTCAGCCACAGGCTACAGGAGCTGAGGCTGAG 720
661 AACGAGGCTGAGACCGCTGCGCAGGCTCAGCCACAGGCTTACAGCAGCTGAGGCTGAG 720
721 TTCGAGGACCCAGGCTGATGACCTGTGGAGCTTGGGAGCTGCGCAGCTCCAGCTCAGGAC 780
721 TTCGAGGACCCAGGCTGATGACCTGTGGAGCTTGGGAGCTTGGGAGCTTGGGAGCTCAGGAC 780
781 AAGGAGCTGAGGCTGTCGGGAGGAGCTCAAGCACTTCGAAAGCCAAATTCGAGAGACAC 840
781 AAGGAGCTGAGGCTGTCGGGAGGAGCTCAAGCACTTCGAAAGCCAAATTCGAGAGACAC 840
841 AACCACTACAGAGCAGCTGAGATTTGGCAGAGAGCTGAGGAGCTGAGGAGAGCGGTG 900
841 AACCACTACAGAGCAGCTGAGATTTGGCAGAGAGCTGAGGAGCTGAGGAGAGCGGTG 900
901 GCGCAGCGGAGGCTGTGAGCGCAGCCGCGAGAGCACTGCTGAGGAGGCGGAGCC 960
901 GCGCAGCGGAGGCTGTGAGCGCAGCCGCGAGAGCACTGCTGAGGAGGCGGAGCC 960
961 AAGGAGCTGAGGCTTACAGGCTGAAGAGCACTTCGAGGAGCTGTCGGGAGGATCTCCAGA 1020
961 AAGGAGCTGAGGCTTACAGGCTGAAGAGCACTTCGAGGAGCTGTCGGGAGGATCTCCAGA 1020
1021 GCTGCGCGAGGACAGAAACCGGTGCAACCCGCGCTCCAGAGCAGTGTGCCCAACAG 1080
1021 GCTGCGCGAGGACAGAAACCGGTGCAACCCGCGCTCCAGAGCAGTGTGCCCAACAG 1080
1081 TCGCAGCTCCCCCAACAGCGCTTTCGATTTGCGCCCTTGACAAGGCCATCACCCAGGAA 1140
1081 TCGCAGCTCCCCCAACAGCGCTTTCGATTTGCGCCCTTGACAAGGCCATCACCCAGGAA 1140
1141 CAGTGCAGGCGCGGCTGCTGCTACATCCCTGCAAAAGCAGGGCTGAGGAGGCCAG 1200
1141 CAGTGCAGGCGCGGCTGCTGCTACATCCCTGCAAAAGCAGGGCTGAGGAGGCCAG 1200
1201 ATGGGCGAGGCTGTGCTTTTCCCAACCGAGCTACCCAGCTACAAGCTGAGAACCTG 1260

1201 ATGGGCGAGCCTGTGCTTTTCCCAACCGAGCTACCCAGCTACAAGCTGAGAACCTG 1260
1261 AGCTCTCTGAAATGGGCTACACGCGCACTGACCGCTACCAACCGCTTCTTCCCC 1320
1261 AGCTCTCTGAAATGGGCTACACGCGCACTGACCGCTACCAACCGCTTCTTCCCC 1320
1321 AAGGACATCTCTGACCTCGCGCTGAGCGTGTGATGATGGAGACTGAGAAACCGCTTCCACTTC 1380
1321 AAGGACATCTCTGACCTCGCGCTGAGCGTGTGATGATGGAGACTGAGAAACCGCTTCCACTTC 1380
1381 AGCATCAAGATCCAGCTAAACAGGCGCTACGAGGTGCCCTTGGAGACCCGCGGTGTCAC 1440
1381 AGCATCAAGATCCAGCTAAACAGGCGCTACGAGGTGCCCTTGGAGACCCGCGGTGTCAC 1440
1441 AGCCGGGACCGCTCCCACTCTACACGCTGAGGTTCCTCCAGAGAGCCCTTCGGGGTGATC 1500
1441 AGCCGGGACCGCTCCCACTCTACACGCTGAGGTTCCTCCAGAGAGCCCTTCGGGGTGATC 1500
1501 GTGCAACCGGAGCTGAGCGCGCTGTCTGTGAACAGAGCGTGGCGCCCTGTCTTCTTT 1560
1501 GTGCAACCGGAGCTGAGCGCGCTGTCTGTGAACAGAGCGTGGCGCCCTGTCTTCTTT 1560
1561 GCGGACAGTTCCTTCAGCTGTCCACTGCTGCTGCCCTGCGAGTATACAGGCGCTCGCC 1620
1561 GCGGACAGTTCCTTCAGCTGTCCACTGCTGCTGCCCTGCGAGTATACAGGCGCTCGCC 1620
1621 GAGCACTCAGTCCCTGATGCTCAGCAGCAGCTGAGACCGCTGAGCAACCGCTGAGAACCGG 1680
1621 GAGCACTCAGTCCCTGATGCTCAGCAGCAGCTGAGACCGCTGAGCAACCGCTGAGAACCGG 1680
1681 GACCTTGGCGCCCAACCGCGGTGCGAACTCTACGGGTCTACCCCTTCTACCTGGCGCTG 1740
1681 GACCTTGGCGCCCAACCGCGGTGCGAACTCTACGGGTCTACCCCTTCTACCTGGCGCTG 1740
1741 GAGGACCGGCTGCGCACAGGCGGTCTCTGCTTAAACAGCAATGCCATGATGTGTC 1800
1741 GAGGACCGGCTGCGCACAGGCGGTCTCTGCTTAAACAGCAATGCCATGATGTGTC 1800
1801 CTGCAAGCGAGCCTGCTTACCTAGCTGAGGTGCAACAGGTGGGATCTGATGATCTACATC 1860
1801 CTGCAAGCGAGCCTGCTTACCTAGCTGAGGTGCAACAGGTGGGATCTGATGATCTACATC 1860
1861 TTCCTGGGCGCAGACCCCAAGAGCGTGTGTCAGCAGTACTCTGAGCTTGTGGGATACCGG 1920
1861 TTCCTGGGCGCAGACCCCAAGAGCGTGTGTCAGCAGTACTCTGAGCTTGTGGGATACCGG 1920
1921 TTCATGCGGCACTACTGGGCGCTGCTTCCACTGTGCGCTGGGCTACTCTCCACC 1980
1921 TTCATGCGGCACTACTGGGCGCTGCTTCCACTGTGCGCTGGGCTACTCTCCACC 1980
1981 GCTATCACCGCGCAGGTGTGGAGAAACATGACAGGCGCCACTTCCCTCGACGTCCTCAA 2040
1981 GCTATCACCGCGCAGGTGTGGAGAAACATGACAGGCGCCACTTCCCTCGACGTCCTCAA 2040
2041 TGGAAACGACTTGACTACATGGAATCCCGGAGGACTTCACTTCAACAAAGATGGCTTC 2100
2041 TGGAAACGACTTGACTACATGGAATCCCGGAGGACTTCACTTCAACAAAGATGGCTTC 2100
2101 CCGGACTTCCCGGCTTCCAGAGTGTGACAGGCGCGCGCTTACATGATGATC 2160
2101 CCGGACTTCCCGGCTTCCAGAGTGTGACAGGCGCGCGCTTACATGATGATC 2160
2161 GTGGATCTCTGCTCAGAGCTGCGGCGCTGCGCGGAGCTACAGGCGCTTACAGAGGCT 2220
2161 GTGGATCTCTGCTCAGAGCTGCGGCGCTGCGCGGAGCTACAGGCGCTTACAGAGGCT 2220
2221 CTGCGGAGGGGGTTTTTTCATCAACAGAGACCGGCGCTGATTTGGGAAGGTATGG 2280
2221 CTGCGGAGGGGGTTTTTTCATCAACAGAGACCGGCGCTGATTTGGGAAGGTATGG 2280
2281 CCGGGCTCCACTGCTTCCCGACTTCCCAACCGAGCTTCCCGCTGGCTGGTGGAGGAC 2340
2281 CCGGGCTCCACTGCTTCCCGACTTCCCAACCGAGCTTCCCGCTGGCTGGTGGAGGAC 2340

|||||
361 CAGGTGACAGCAACTCCCTCAGTGGCACCCAGGAAGACGGGTGGATGACCCAGGCTG 420
QY
421 GAAGAAGCTGGACAGCAAGCGGAGACCTCTGGGAATTTCTCGGCGGAAGAACTGGACAG 480
DB
421 GAAGAAGCTGGACAGCAAGCGGAGACCTCTGGGAATTTCTCGGCGGAAGAACTGGACAG 480
QY
481 CTCTGGCGGAGTTCTCTGCATCACAAGAGAAAGTTACAGAGTACAACTGCTCTGTGGAG 540
DB
481 CTCTGGCGGAGTTCTCTGCATCACAAGAGAAAGTTACAGAGTACAACTGCTCTGTGGAG 540
QY
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAAACGTCATTAGCCCTCGGACCTGAGCGAC 600
DB
541 ACCCTGAGCAGGACCGAAGAAATCCACGAGAAACGTCATTAGCCCTCGGACCTGAGCGAC 600
QY
601 ATCAAGGCGAGGCTCTGCACAGCAGGACACGAGGCTGAAGGAGAGCTGCGCAGCATC 660
DB
601 ATCAAGGCGAGGCTCTGCACAGCAGGACACGAGGCTGAAGGAGAGCTGCGCAGCATC 660
QY
661 AACCAAGGCGCTGGACCGCTCGCAGGAGTCAGCCACAGGCTACAGCACTGAGGCTGAG 720
DB
661 AACCAAGGCGCTGGACCGCTCGCAGGAGTCAGCCACAGGCTACAGCACTGAGGCTGAG 720
QY
721 TTGAGGAGCCGACAGGAGTGAATGACTGTGGGACCTGGGCGAGTCCGCCAACTCAGCGAC 780
DB
721 TTGAGGAGCCGACAGGAGTGAATGACTGTGGGACCTGGGCGAGTCCGCCAACTCAGCGAC 780
QY
781 AAGGAGCTGGAGGCTTCCGGAGGAGCTCAAGCACTTCGAGGCCAAATCAGAGACAC 840
DB
781 AAGGAGCTGGAGGCTTCCGGAGGAGCTCAAGCACTTCGAGGCCAAATCAGAGACAC 840
QY
841 AACCACTACAGAGCAGCTGGAGATTGGCGCACGAGAACTGAGGACACGACAGAGCGCTG 900
DB
841 AACCACTACAGAGCAGCTGGAGATTGGCGCACGAGAACTGAGGACACGACAGAGCGCTG 900
QY
901 GCGCAGCGGAGCTGAGCCGCTGAGCCGAGCGCAGAGACGACCCCTGTGAGGGGCGGACC 960
DB
901 GCGCAGCGGAGCTGAGCCGCTGAGCCGAGCGCAGAGACGACCCCTGTGAGGGGCGGACC 960
QY
961 AAGGAGCTGGGCTACACGCTGAGAGCATCTGCGAGGAGCTGCGAGGAGTCTCCAGA 1020
DB
961 AAGGAGCTGGGCTACACGCTGAGAGCATCTGCGAGGAGCTGCGAGGAGTCTCCAGA 1020
QY
1021 GCTCGCGCGAGCAGAAACCGGTGACACCCCGCGCTCCAGAGCAGTGCACACAG 1080
DB
1021 GCTCGCGCGAGCAGAAACCGGTGACACCCCGCGCTCCAGAGCAGTGCACACAG 1080
QY
1081 TCGCAGCTCCCCCMAAGCCGCTTCGATTGGGCCCTTGACAGGCCATCACCCAGGAA 1140
DB
1081 TCGCAGCTCCCCCMAAGCCGCTTCGATTGGGCCCTTGACAGGCCATCACCCAGGAA 1140
QY
1141 CAGTGGAGGCGCGGCTGCTGTACATCCCTGCAAGCAGGGGTGAGGAGGCCAG 1200
DB
1141 CAGTGGAGGCGCGGCTGCTGTACATCCCTGCAAGCAGGGGTGAGGAGGCCAG 1200
QY
1201 ATGGGCGAGCCCTGTGTCTTTCCACCCAGCTACCCAGCTACAAAGCTGGAGAACCTG 1260
DB
1201 ATGGGCGAGCCCTGTGTCTTTCCACCCAGCTACCCAGCTACAAAGCTGGAGAACCTG 1260
QY
1261 AGCTCTCTGAATGGGCTACAGGCGCACCTGACCCGTTACACCCGACCTTCTCCCC 1320
DB
1261 AGCTCTCTGAATGGGCTACAGGCGCACCTGACCCGTTACACCCGACCTTCTCCCC 1320
QY
1321 AAGGACATCTGACCTGGGCTGGAGCTGATGATGGAGACTGAGAACCGCTCCACTTC 1380
DB
1321 AAGGACATCTGACCTGGGCTGGAGCTGATGATGGAGACTGAGAACCGCTCCACTTC 1380
QY
1381 ACGATCAAAAGATCCAGCTAACAGGGGCTACGAGGTGCCCTTGGAGACCCCGGCTGCCAC 1440
DB
1381 ACGATCAAAAGATCCAGCTAACAGGGGCTACGAGGTGCCCTTGGAGACCCCGGCTGCCAC 1440
QY
1441 AGCCGGGACCGTCCCACTTACAGCTGGAGTTCTCCGAGGAGCCCTTCGGGCTGATC 1500
|||||

1441 AGCCGGGACCGTCCCACTCTTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATC 1500
QY
1501 GTGCACCCGCGAGCTGACGCGCGCTGTCTGTGAACAACGACGCTGCGGCCCTGTCTTT 1560
DB
1501 GTGCACCCGCGAGCTGACGCGCGCTGTCTGTGAACAACGACGCTGCGGCCCTGTCTTT 1560
QY
1561 GCGGACCAAGTTCTTTTTCAGCTGTCCACTCGCTGCTCGCTGCTGCTGCTGCTGCTGCT 1620
DB
1561 GCGGACCAAGTTCTTTTTCAGCTGTCCACTCGCTGCTCGCTGCTGCTGCTGCTGCTGCT 1620
QY
1621 GAGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB
1621 GAGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
QY
1681 GACCTTGGCGCCGACGCGCTGCGAACTCTACGGGTCTCACCTTTCTTACCTGCGCTG 1740
DB
1681 GACCTTGGCGCCGACGCGCTGCGAACTCTACGGGTCTCACCTTTCTTACCTGCGCTG 1740
QY
1741 GAGGACGCGGCTGCGGACACGCGGTGTTCTGTCTTAAACAGCAATGCCATGATGTGTC 1800
DB
1741 GAGGACGCGGCTGCGGACACGCGGTGTTCTGTCTTAAACAGCAATGCCATGATGTGTC 1800
QY
1801 CTGACGCGGAGCCCTGCTTCTGCTGAGGCTGACAGGCTGAGGATCCTGGATGCTTACATC 1860
DB
1801 CTGACGCGGAGCCCTGCTTCTGCTGAGGCTGACAGGCTGAGGATCCTGGATGCTTACATC 1860
QY
1861 TTCTTGGGCGCAGAGCCCAAGAGCGTGTGACGAGTACCTGAGCTGTTGTGGGATACCG 1920
DB
1861 TTCTTGGGCGCAGAGCCCAAGAGCGTGTGACGAGTACCTGAGCTGTTGTGGGATACCG 1920
QY
1921 TTCTTGGGCGCAGTCTGCGGCTTCCACCTGTGCGCTGCGGCTGCTTCTTCTCCACC 1980
DB
1921 TTCTTGGGCGCAGTCTGCGGCTTCCACCTGTGCGCTGCGGCTGCTTCTTCTCCACC 1980
QY
1981 GCTATCACCCGCGAGTGTGGAGAACATGACAGGCGCCACTTCCCTTGGAGCTCCAA 2040
DB
1981 GCTATCACCCGCGAGTGTGGAGAACATGACAGGCGCCACTTCCCTTGGAGCTCCAA 2040
QY
2041 TGGAAAGACCTTGGACTACATGAGCTCCGCGGAGGACTTTCAGCTTCAACAGAGATGGCTTC 2100
DB
2041 TGGAAAGACCTTGGACTACATGAGCTCCGCGGAGGACTTTCAGCTTCAACAGAGATGGCTTC 2100
QY
2101 CGGAGCTTCCGCGCATGCTGAGGCTGACACAGGCGGCGCGCTTACATGATGATC 2160
DB
2101 CGGAGCTTCCGCGCATGCTGAGGCTGACACAGGCGGCGCGCTTACATGATGATC 2160
QY
2161 GTGATCTCTGCTCAGCAGCTCGGCGCTCGCGGAGCTACAGGCTTACGACGAGGT 2220
DB
2161 GTGATCTCTGCTCAGCAGCTCGGCGCTCGCGGAGCTACAGGCTTACGACGAGGT 2220
QY
2221 CTGCGGAGGCGGTTTTCATCACCAAGAGACCGGCGCAGCTGATTTGGGAGGATGG 2280
DB
2221 CTGCGGAGGCGGTTTTCATCACCAAGAGACCGGCGCAGCTGATTTGGGAGGATGG 2280
QY
2281 CCGGCTTCCGCTTCCCGACTTCCCAACCCACAGGCTGCGCTTGGGAGGAC 2340
DB
2281 CCGGCTTCCGCTTCCCGACTTCCCAACCCACAGGCTGCGCTTGGGAGGAC 2340
QY
2341 ATGGTGGCTGAGTTCCATGACAGGCTTCCGAGGCTTGGATGATGATGATGATGATGAT 2400
DB
2341 ATGGTGGCTGAGTTCCATGACAGGCTTCCGAGGCTTGGATGATGATGATGATGATGAT 2400
QY
2401 CTTTCAAACTTCTCAGAGGCTCTGAGGAGGCTGCGGCTGCGGCTGCGGAGAACCCCA 2460
DB
2401 CTTTCAAACTTCTCAGAGGCTCTGAGGAGGCTGCGGCTGCGGCTGCGGAGAACCCCA 2460
QY
2461 CTTTCAAACTTCTCAGAGGCTCTGAGGAGGCTGCGGCTGCGGCTGCGGAGAACCCCA 2520
DB
2461 CTTTCAAACTTCTCAGAGGCTCTGAGGAGGCTGCGGCTGCGGCTGCGGAGAACCCCA 2520
QY
2521 CACCAAGTTTCTCTCAGAGGCTCTGAGGAGGCTGCGGCTGCGGCTGCGGAGAACCCCA 2580
DB
2521 CACCAAGTTTCTCTCAGAGGCTCTGAGGAGGCTGCGGCTGCGGCTGCGGAGAACCCCA 2580

Qy	2581	GCCTCCCA	CAGGGCGCTGTGTAAAGGCTCGGGGACA	CGCCCAATTGTGATCTCCCGCTCG	2640
Db	2581	GCCTCCCA	CAGGGCGCTGTGTAAAGGCTCGGGGACA	CGCCCAATTGTGTGATCTCCCGCTCG	2640
Qy	2641	ACCTTTGCT	GGCCACAGCCGATACGCGGCCACTGGACGGGGACG	TGTGGAGCTCTCTGG	2700
Db	2641	ACCTTTGCT	GGCCACAGCCGATACGCGGCCACTGGACGGGGACG	TGTGGAGCTCTCTGG	2700
Qy	2701	GAGCAGCT	CGCCTCTCTCGTGCCAGAAATCCTGCAGTTAACTCTGGGGGTGCCTCTG	2760	
Db	2701	GAGCAGCT	CGCCTCTCTCGTGCCAGAAATCCTGCAGTTAACTCTGGGGGTGCCTCTG	2760	
Qy	2761	GTCGGGGCC	GACGCTCTGGGCTTCTCTGGGCAACACCTCAGAGAGCTGTGTGTGCGCTGG	2820	
Db	2761	GTCGGGGCC	GACGCTCTGGGCTTCTCTGGGCAACACCTCAGAGAGCTGTGTGTGCGCTGG	2820	
Qy	2821	ACCCAGCT	GGGGGCTTCTACCCCTTCATGCGGAACCAACACGCTGCTCAGCTTGCCCC	2880	
Db	2821	ACCCAGCT	GGGGGCTTCTACCCCTTCATGCGGAACCAACACGCTGCTCAGCTTGCCCC	2880	
Qy	2881	CAGGAGCG	TACAGCTTCAGCGAGCCGGCCAGCAGGCCATGAGGAAGGCCCTCACCCCTG	2940	
Db	2881	CAGGAGCG	TACAGCTTCAGCGAGCCGGCCAGCAGGCCATGAGGAAGGCCCTCACCCCTG	2940	
Qy	2941	CGCTACG	CACTCTCCCCCACTCTACACACTGTTTCCACGAGCCACGTCGCGGGGAG	3000	
Db	2941	CGCTACG	CACTCTCCCCCACTCTACACACTGTTTCCACGAGCCACGTCGCGGGGAG	3000	
Qy	3001	ACCGTGCC	CGGCGCCCTCTTCCTGGAGTTCCCAAGGACTCTAGCACTGGACTGTGGAC	3060	
Db	3001	ACCGTGCC	CGGCGCCCTCTTCCTGGAGTTCCCAAGGACTCTAGCACTGGACTGTGGAC	3060	
Qy	3061	CACCAGCT	CTCTGTGGGGGAGGCCCTGCTATACACCCAGTGCTCAGAGCCGGGAGGCC	3120	
Db	3061	CACCAGCT	CTCTGTGGGGGAGGCCCTGCTATACACCCAGTGCTCAGAGCCGGGAGGCC	3120	
Qy	3121	GAAGTGAC	TGGCTACTTCTCCCTTTGGGCACATGTCAGACCTGACAGCGGTGCCAATAGAG	3180	
Db	3121	GAAGTGAC	TGGCTACTTCTCCCTTTGGGCACATGTCAGACCTGACAGCGGTGCCAATAGAG	3180	
Qy	3181	GCCCTTGG	CAGCTTCCACCCCACTGTGAGCTTCCCGTGAGCCAGGCCATTCACAGCGAG	3240	
Db	3181	GCCCTTGG	CAGCTTCCACCCCACTGTGAGCTTCCCGTGAGCCAGGCCATTCACAGCGAG	3240	
Qy	3241	GGGCAGTG	GGGTGACGCTGCGGGCCCCCTTGACACCATCAAGTTCACCTCCGGGGTGGG	3300	
Db	3241	GGGCAGTG	GGGTGACGCTGCGGGCCCCCTTGACACCATCAAGTTCACCTCCGGGGTGGG	3300	
Qy	3301	TACATCAT	CCCCCTGACGGGCCCTTGGCTTCACAAACACAGAGTCCCCGCCAGCAGCCCATG	3360	
Db	3301	TACATCAT	CCCCCTGACGGGCCCTTGGCTTCACAAACACAGAGTCCCCGCCAGCAGCCCATG	3360	
Qy	3361	GCCCTGCT	GTGGCCCTTAAACAAGGGTGGAGAGCCCGAGGGGAGCTGTTCTGGGACGAT	3420	
Db	3361	GCCCTGCT	GTGGCCCTTAAACAAGGGTGGAGAGCCCGAGGGGAGCTGTTCTGGGACGAT	3420	
Qy	3421	GGAGAGCT	TGGAAAGTCTGGAGCGAGGGGCCCTTACACAGGTCATCTTCTTGGGCCAGG	3480	
Db	3421	GGAGAGCT	TGGAAAGTCTGGAGCGAGGGGCCCTTACACAGGTCATCTTCTTGGGCCAGG	3480	
Qy	3481	AATAACAC	GCATGTGTAAGTGTGACCATGAGGAGCTGCCCTGACAGCTG	3540	
Db	3481	AATAACAC	GCATGTGTAAGTGTGACCATGAGGAGCTGCCCTGACAGCTG	3540	
Qy	3541	CAGAAGGT	GATCTGTCTGGGCGTGGCCACCGCGCCCGCAGAGGTCCTTCTCCAAACGGGTGC	3600	
Db	3541	CAGAAGGT	GATCTGTCTGGGCGTGGCCACCGCGCCCGCAGAGGTCCTTCTCCAAACGGGTGC	3600	
Qy	3601	CCTGTCTC	CAACTTCACTCAGCCCCGACACCAAGGTCCTGACATCTGTGTCTGCTG	3660	
Db	3601	CCTGTCTC	CAACTTCACTCAGCCCCGACACCAAGGTCCTGACATCTGTGTCTGCTG	3660	

```

QY      3661  TTGATGGAGACAGATTTCCTGCAGCTGGTGTACTCGAG 3702
          |||||
Db      3661  TTGATGGAGACAGATTTCCTGCAGCTGGTGTACTCGAG 3702
          |||||

RESULT 3
US-10-761-530-3
; Sequence 3, Application US/10761530
; Publication No. US20040248262A1
; GENERAL INFORMATION:
; APPLICANT: Koeberl, Dwight D.
; APPLICANT: Sun, Baodong
; TITLE OF INVENTION: IMPROVED CONSTRUCTS FOR EXPRESSING LYOSOMAL POLYPEPTIDES
; FILE REFERENCE: 5405-280
; CURRENT APPLICATION NUMBER: US/10/761,530
; CURRENT FILING DATE: 2004-01-21
; PRIOR APPLICATION NUMBER: US 60/441,789
; PRIOR FILING DATE: 2003-01-22
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 3
; LENGTH: 3026
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-761-530-3

Query Match           71.5%; Score 2647.8; DB 18; Length 3026;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2652; Conservative       7; Indels    0; Gaps    0;

```

Qy 3661 TTGATGGGAGAGCAGTTTCTCGTCAGCTGGTGTGACTCGAG 3702
 |||||
 Db 3661 TTGATGGGAGAGCAGTTTCTCGTCAGCTGGTGTGACTCGAG 3702
 |||||

1696 CCCTGTCGAACTCTACGGGTCTCAACCTTTCTACCTGCGCTGGAGACGGCGGGTCG 1755
1697 |
1698 |
1699 |
1700 |
1701 |
1702 |
1703 |
1704 |
1705 |
1706 |
1707 |
1708 |
1709 |
1710 |
1711 |
1712 |
1713 |
1714 |
1715 |
1716 |
1717 |
1718 |
1719 |
1720 |
1721 |
1722 |
1723 |
1724 |
1725 |
1726 |
1727 |
1728 |
1729 |
1730 |
1731 |
1732 |
1733 |
1734 |
1735 |
1736 |
1737 |
1738 |
1739 |
1740 |
1741 |
1742 |
1743 |
1744 |
1745 |
1746 |
1747 |
1748 |
1749 |
1750 |
1751 |
1752 |
1753 |
1754 |
1755 |
1756 |
1757 |
1758 |
1759 |
1760 |
1761 |
1762 |
1763 |
1764 |
1765 |
1766 |
1767 |
1768 |
1769 |
1770 |
1771 |
1772 |
1773 |
1774 |
1775 |
1776 |
1777 |
1778 |
1779 |
1780 |
1781 |
1782 |
1783 |
1784 |
1785 |
1786 |
1787 |
1788 |
1789 |
1790 |
1791 |
1792 |
1793 |
1794 |
1795 |
1796 |
1797 |
1798 |
1799 |
1800 |
1801 |
1802 |
1803 |
1804 |
1805 |
1806 |
1807 |
1808 |
1809 |
1810 |
1811 |
1812 |
1813 |
1814 |
1815 |
1816 |
1817 |
1818 |
1819 |
1820 |
1821 |
1822 |
1823 |
1824 |
1825 |
1826 |
1827 |
1828 |
1829 |
1830 |
1831 |
1832 |
1833 |
1834 |
1835 |
1836 |
1837 |
1838 |
1839 |
1840 |
1841 |
1842 |
1843 |
1844 |
1845 |
1846 |
1847 |
1848 |
1849 |
1850 |
1851 |
1852 |
1853 |
1854 |
1855 |
1856 |
1857 |
1858 |
1859 |
1860 |
1861 |
1862 |
1863 |
1864 |
1865 |
1866 |
1867 |
1868 |
1869 |
1870 |
1871 |
1872 |
1873 |
1874 |
1875 |
1876 |
1877 |
1878 |
1879 |
1880 |
1881 |
1882 |
1883 |
1884 |
1885 |
1886 |
1887 |
1888 |
1889 |
1890 |
1891 |
1892 |
1893 |
1894 |
1895 |
1896 |
1897 |
1898 |
1899 |
1900 |
1901 |
1902 |
1903 |
1904 |
1905 |
1906 |
1907 |
1908 |
1909 |
1910 |
1911 |
1912 |
1913 |
1914 |
1915 |
1916 |
1917 |
1918 |
1919 |
1920 |
1921 |
1922 |
1923 |
1924 |
1925 |
1926 |
1927 |
1928 |
1929 |
1930 |
1931 |
1932 |
1933 |
1934 |
1935 |
1936 |
1937 |
1938 |
1939 |
1940 |
1941 |
1942 |
1943 |
1944 |
1945 |
1946 |
1947 |
1948 |
1949 |
1950 |
1951 |
1952 |
1953 |
1954 |
1955 |
1956 |
1957 |
1958 |
1959 |
1960 |
1961 |
1962 |
1963 |
1964 |
1965 |
1966 |
1967 |
1968 |
1969 |
1970 |
1971 |
1972 |
1973 |
1974 |
1975 |
1976 |
1977 |
1978 |
1979 |
1980 |
1981 |
1982 |
1983 |
1984 |
1985 |
1986 |
1987 |
1988 |
1989 |
1990 |
1991 |
1992 |
1993 |
1994 |
1995 |
1996 |
1997 |
1998 |
1999 |
2000 |

2776 TCGCGCTTCTCTGGGCAACACCTCAGAGGAGCTGTGTGTGCTGGACCCAGCTGGGGGCC 2835
2777 |
2778 |
2779 |
2780 |
2781 |
2782 |
2783 |
2784 |
2785 |
2786 |
2787 |
2788 |
2789 |
2790 |
2791 |
2792 |
2793 |
2794 |
2795 |
2796 |
2797 |
2798 |
2799 |
2800 |
2801 |
2802 |
2803 |
2804 |
2805 |
2806 |
2807 |
2808 |
2809 |
2810 |
2811 |
2812 |
2813 |
2814 |
2815 |
2816 |
2817 |
2818 |
2819 |
2820 |
2821 |
2822 |
2823 |
2824 |
2825 |
2826 |
2827 |
2828 |
2829 |
2830 |
2831 |
2832 |
2833 |
2834 |
2835 |
2836 |
2837 |
2838 |
2839 |
2840 |
2841 |
2842 |
2843 |
2844 |
2845 |
2846 |
2847 |
2848 |
2849 |
2850 |
2851 |
2852 |
2853 |
2854 |
2855 |
2856 |
2857 |
2858 |
2859 |
2860 |
2861 |
2862 |
2863 |
2864 |
2865 |
2866 |
2867 |
2868 |
2869 |
2870 |
2871 |
2872 |
2873 |
2874 |
2875 |
2876 |
2877 |
2878 |
2879 |
2880 |
2881 |
2882 |
2883 |
2884 |
2885 |
2886 |
2887 |
2888 |
2889 |
2890 |
2891 |
2892 |
2893 |
2894 |
2895 |
2896 |
2897 |
2898 |
2899 |
2900 |
2901 |
2902 |
2903 |
2904 |
2905 |
2906 |
2907 |
2908 |
2909 |
2910 |
2911 |
2912 |
2913 |
2914 |
2915 |
2916 |
2917 |
2918 |
2919 |
2920 |
2921 |
2922 |
2923 |
2924 |
2925 |
2926 |
2927 |
2928 |
2929 |
2930 |
2931 |
2932 |
2933 |
2934 |
2935 |
2936 |
2937 |
2938 |
2939 |
2940 |
2941 |
2942 |
2943 |
2944 |
2945 |
2946 |
2947 |
2948 |
2949 |
2950 |
2951 |
2952 |
2953 |
2954 |
2955 |
2956 |
2957 |
2958 |
2959 |
2960 |
2961 |
2962 |
2963 |
2964 |
2965 |
2966 |
2967 |
2968 |
2969 |
2970 |
2971 |
2972 |
2973 |
2974 |
2975 |
2976 |
2977 |
2978 |
2979 |
2980 |
2981 |
2982 |
2983 |
2984 |
2985 |
2986 |
2987 |
2988 |
2989 |
2990 |
2991 |
2992 |
2993 |
2994 |
2995 |
2996 |
2997 |
2998 |
2999 |
3000 |

RESULT 4

US-10-761-530-1
; Sequence 1, Application US/10761530
; Publication No. US2004024826A1
; GENERAL INFORMATION:
; APPLICANT: Koerber, Dwight D.
; APPLICANT: Sun, Baodong
; TITLE OF INVENTION: IMPROVED CONSTRUCTS FOR EXPRESSING LYOSOMAL POLYPEPTIDES

FILE REFERENCE: 5405-280
CURRENT APPLICATION NUMBER: US/10/761,530
CURRENT FILING DATE: 2004-01-21
PRIOR APPLICATION NUMBER: US 60/441,789
PRIOR FILING DATE: 2003-01-22
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 3846
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (442)..(3300)
US-10-761-530-1

Query Match 71.5%; Score 2647.8; DB 18; Length 3846;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2652; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY	1036	GAACCGGTGCACACCCCGCGCTCCAGAGCAGTGCCACACAGTGCAGCTGCCCGCC	1095
DB	640	GATGCCCGGACACCCCGCGCTCCAGAGCAGTGCCACACAGTGCAGCTGCCCGCC	699
QY	1096	AACAGCGCTTCGATTGCGCCCTGACAAAGCCATCACCCAGAAACAGTGCAGGCGCCG	1155
DB	700	AACAGCGCTTCGATTGCGCCCTGACAAAGCCATCACCCAGAAACAGTGCAGGCGCCG	759
QY	1156	GGTGTCTACATCTCTGAAAGCAGGCGCTCAGGAGCCACAGTGGGCGAGCCCTGG	1215
DB	760	GGTGTCTACATCTCTGAAAGCAGGCGCTCAGGAGCCACAGTGGGCGAGCCCTGG	819
QY	1216	TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGAGAACCTGAGCTCTCTGAAATG	1275
DB	820	TGCTTCTTCCACCCAGCTACCCAGCTACAAGCTGAGAACCTGAGCTCTCTGAAATG	879
QY	1276	GGCTACACGGCCACCTGACCCCGTACACCCCACTTCTTCCCAAGGACATCTTGACC	1335
DB	880	GGCTACACGGCCACCTGACCCCGTACACCCCACTTCTTCCCAAGGACATCTTGACC	939
QY	1336	CTCGGCTGACGTGATGATGAGACTGAGAAACCGGCTCACTTACGATCAAGATCCA	1395
DB	940	CTCGGCTGACGTGATGATGAGACTGAGAAACCGGCTCACTTACGATCAAGATCCA	999
QY	1396	GCTAACAGGCGCTACGAGTGCCCTTGGAGACCCCGCTGTCCACAGCGGGCACCGTCC	1455
DB	1000	GCTAACAGGCGCTACGAGTGCCCTTGGAGACCCCGCTGTCCACAGCGGGCACCGTCC	1059
QY	1456	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGCAACCGGCACTG	1515
DB	1060	CCACTCTACAGCGTGGAGTTCTCCGAGGAGCCCTTCGGGGTGATCGTGCAACCGGCACTG	1119
QY	1516	GACGGCGCGTGTGTGAACACAGACGGTGGCGCCCTGTCTTTGGCGAACAGATTCCTT	1575
DB	1120	GACGGCGCGTGTGTGAACACAGACGGTGGCGCCCTGTCTTTGGCGAACAGATTCCTT	1179
QY	1576	CAGCTGTCAACCTCGCTGCGCTGCGAGTATACAGGCTCGCGAGCACCTCAGTCCC	1635
DB	1180	CAGCTGTCAACCTCGCTGCGCTGCGAGTATACAGGCTCGCGAGCACCTCAGTCCC	1239
QY	1636	CTGATGTACAGCACCACTGAGACAGGATCACCTGTGGAAACCGGACCTTGGCGCCACG	1695
DB	1240	CTGATGTACAGCACCACTGAGACAGGATCACCTGTGGAAACCGGACCTTGGCGCCACG	1299
QY	1696	CCCGGTGCGAACCTCTACGGGTCTCACCTTTTACCTGGCGCTGAGAGACGGCGGGTCG	1755
DB	1300	CCCGGTGCGAACCTCTACGGGTCTCACCTTTTACCTGGCGCTGAGAGACGGCGGGTCG	1359
QY	1756	GCACAGGGGTGTCTCTTAAACAGCAATGCCATGATGTGTCTGACAGCGGCGCT	1815
DB	1360	GCACAGGGGTGTCTCTTAAACAGCAATGCCATGATGTGTCTGACAGCGGCGCT	1419
QY	1816	GCCCTTAGCTGGAGTGCACAGGTGGGATCTGTGATGTCTTACATCTTCTCGGGCCACAG	1875

DB	1420	GGCCTTAGCTGGAGTGCACAGGTGGATCTCTGGATGTCTACATCTTCTTGGGCCACAG	1479
QY	1876	CCCAAGAGCGTGTGTGACAGTACCTTGGACGTTGTGGGATACCCGTTTATGCCGCATAC	1935
DB	1480	CCCAAGAGCGTGTGTGACAGTACCTTGGACGTTGTGGGATACCCGTTTATGCCGCATAC	1539
QY	1936	TGGGGCTGGGCTTCCACTGTGCGCTGGGGTACTCTCTCACCGCTATACCCGCCAG	1995
DB	1540	TGGGGCTGGGCTTCCACTGTGCGCTGGGGTACTCTCTCACCGCTATACCCGCCAG	1599
QY	1996	GTGGTGGAGAACATGACACAGGCGCCACTTCCCGCTGGAGCTCAATGGAAACGACCTGGAC	2055
DB	1600	GTGGTGGAGAACATGACACAGGCGCCACTTCCCGCTGGAGCTCAATGGAAACGACCTGGAC	1659
QY	2056	TACATGGACTCCCGGAGGACTTCACTGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCC	2115
DB	1660	TACATGGACTCCCGGAGGACTTCACTGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCC	1719
QY	2116	ATGGTGCAGAGCTGCACACAGGCGCGCTACATGATGATGATGATGATGATGATGATGATG	2175
DB	1720	ATGGTGCAGAGCTGCACACAGGCGCGCTACATGATGATGATGATGATGATGATGATGATG	1779
QY	2176	AGCAGCTCGGGCCCTGCGGGAGCTACAGGCGCTACAGAGGCTCTCGGAGGGGGTT	2235
DB	1780	AGCAGCTCGGGCCCTGCGGGAGCTACAGGCGCTACAGAGGCTCTCGGAGGGGGTT	1839
QY	2236	TTCATCACCAACGAGACCGGCGCTGATTTGGAAAGTATGGCCCGGCTCCACTGGCC	2295
DB	1840	TTCATCACCAACGAGACCGGCGCTGATTTGGAAAGTATGGCCCGGCTCCACTGGCC	1899
QY	2296	TTCCCGGCTTCAACCAACCGGCGCTGATTTGGAAAGTATGGGGGATGATGATGATGATG	2355
DB	1900	TTCCCGGCTTCAACCAACCGGCGCTGATTTGGAAAGTATGGGGGATGATGATGATGATG	1959
QY	2356	CATGACAGGTGCTTCCAGCGCTTGTGATTTGATGATGATGATGATGATGATGATGATG	2415
DB	1960	CATGACAGGTGCTTCCAGCGCTTGTGATTTGATGATGATGATGATGATGATGATGATG	2019
QY	2416	AGAGGCTCTGAGAGCGGCTGCCCAACATAGCTGGAGAACCCACCTACGTCGCTGGG	2475
DB	2020	AGAGGCTCTGAGAGCGGCTGCCCAACATAGCTGGAGAACCCACCTACGTCGCTGGG	2079
QY	2476	GTGGTTGGGGGACCCCTCCAGCGGCGCCATCTGTGCTTCCAGCAACAGATTTCTTCC	2535
DB	2080	GTGGTTGGGGGACCCCTCCAGCGGCGCCATCTGTGCTTCCAGCAACAGATTTCTTCC	2139
QY	2536	ACACACTACACCTTGACAACTCTAGGCGCTGACCGAGCCATCGCTCCACAGGGCG	2595
DB	2140	ACACACTACACCTTGACAACTCTAGGCGCTGACCGAGCCATCGCTCCACAGGGCG	2199
QY	2596	CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGTGGCCAC	2655
DB	2200	CTGGTGAAGGCTCGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGTGGCCAC	2259
QY	2656	GGCGGATACCGCGCCACTGGAGCGGGGACGTGTGGAGTCTCTGGAGACAGCTCGGCTCC	2715
DB	2260	GGCGGATACCGCGCCACTGGAGCGGGGACGTGTGGAGTCTCTGGAGACAGCTCGGCTCC	2319
QY	2716	TCCGTGCCAGAAATCTCGAGTTTAACTTGTGGGGGTGCTTGTGGGGGCGGAGCTC	2775
DB	2320	TCCGTGCCAGAAATCTCGAGTTTAACTTGTGGGGGTGCTTGTGGGGGCGGAGCTC	2379
QY	2776	TGGCGCTTCTTGGGCAACCTTCAGAGGAGCTGTGTGGCTGGACCCAGCTGGGGGCC	2835
DB	2380	TGGCGCTTCTTGGGCAACCTTCAGAGGAGCTGTGTGTGGCTGGACCCAGCTGGGGGCC	2439
QY	2836	TTCTACCCCTTCTATGCGGAAACCAACAGCTGCTCAGTGTGCCCGCAGGAGCCGTACAG	2895
DB	2440	TTCTACCCCTTCTATGCGGAAACCAACAGCTGCTCAGTGTGCCCGCAGGAGCCGTACAG	2499
QY	2896	TTCAGCGAGCGGCGCCAGAGGCGCATGAGGAAGGCCCTTACCCCTCGGCTAGCGACTCTC	2955

Db 2500 TTCAGCGAGCGCGCCAGCAGGCGCATGAGGAGGCGCTCACCTCGCGCTACGCACTCTCTC 2559
QY CCCCACCTCTACACACTGTTTCCACAGGCGCCAGCTCGCGGGGAGACCGTGGCGCGGCC 3015
Db CCCCACCTCTACACACTGTTTCCACAGGCGCCAGCTCGCGGGGAGACCGTGGCGCGGCC 2619
QY CTCCTTCTGAGGTTCCCAAGGACTTAGCACTGAGCTGTGACCAACAGCTCTCTGTGG 3075
Db CTCCTTCTGAGGTTCCCAAGGACTTAGCACTGAGCTGTGACCAACAGCTCTCTGTGG 2679
QY GGGGAGCGCTCTCATACCCAGGCTTCCAGGCGGGAAGCGCAAGTGAAGTGAAGTGAAG 3135
Db GGGGAGCGCTCTCATACCCAGGCTTCCAGGCGGGAAGCGCAAGTGAAGTGAAGTGAAG 2739
QY TTCCCTCTGGGACATAGTGAAGCTTCCAGGCGGGAAGCGCAAGTGAAGTGAAGTGAAG 3195
Db TTCCCTCTGGGACATAGTGAAGCTTCCAGGCGGGAAGCGCAAGTGAAGTGAAGTGAAG 2799
QY CACCCCTCTGAGCTGAGCTCCCGTGAAGCGGATCCACAGGAGGGGAGTGGGTGAGCG 3255
Db CACCCCTCTGAGCTGAGCTCCCGTGAAGCGGATCCACAGGAGGGGAGTGGGTGAGCG 2859
QY CTGCGGCGCGCTTGACACCATCAAGCTTCCAGCTCCAGCTCCGCGGCTGGGTGAGCG 3315
Db CTGCGGCGCGCTTGACACCATCAAGCTTCCAGCTCCAGCTCCGCGGCTGGGTGAGCG 2919
QY CAGGCGCTTGGCTTCAACACAGAGTCCCGCAGAGCGCCATGCGCTGGTGGCC 3375
Db CAGGCGCTTGGCTTCAACACAGAGTCCCGCAGAGCGCCATGCGCTGGTGGCC 2979
QY CTAACCAAGGTTGAGAGCGCGGAGCTGTTCTGGGACGATGAGAGAGCTTGGAA 3435
Db CTGACCAAGGTTGAGAGCGCGGAGCTGTTCTGGGACGATGAGAGAGCTTGGAA 3039
QY GTGCTGGAGGAGGCGCTTACACAGGTCATCTTCTGCGGAGGATTAACAGATCGTG 3495
Db GTGCTGGAGGAGGCGCTTACACAGGTCATCTTCTGCGGAGGATTAACAGATCGTG 3099
QY AATGAGCTGTGAGTGTGACAGGAGCTGCGCTGCGAGTGCAGAGGAGTGAAGTGAAGT 3555
Db AATGAGCTGTGAGTGTGACAGGAGCTGCGCTGCGAGTGCAGAGGAGTGAAGTGAAGT 3159
QY CTGGGCGTGGCCAGCGCGCCAGCAGGTCCTTCCAAAGGTCCTCTGCTTCCAACTTC 3615
Db CTGGGCGTGGCCAGCGCGCCAGCAGGTCCTTCCAAAGGTCCTCTGCTTCCAACTTC 3219
QY ACCTAGAGCGCGACAGGAGTCTTGGACATCTGTCTCGCTGTTGATGGAGAGCAG 3675
Db ACCTAGAGCGCGACAGGAGTCTTGGACATCTGTCTCGCTGTTGATGGAGAGCAG 3279
QY TTTCTCGTCAGCTGGTGTT 3694
Db TTTCTCGTCAGCTGGTGTT 3298

RESULT 5

US-10-874-242-38
; Sequence 38, Application US/10874242
; Publication No. US20050026252A1
; GENERAL INFORMATION:
; APPLICANT: ESTES, SCOTT
; APPLICANT: ZHANG, WEIQUN
; TITLE OF INVENTION: NOVEL PROMOTERS AND USES THEREOF
; FILE REFERENCE: 07680.0027-00000
; CURRENT APPLICATION NUMBER: US/10/874,242
; CURRENT FILING DATE: 2004-06-24
; PRIOR APPLICATION NUMBER: 60/480,768
; PRIOR FILING DATE: 2003-06-24
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 2859
; TYPE: DNA

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: cDNA encoding alpha-glucosidase.
US-10-874-242-38

Query Match 71.4%; Score 2643.4; DB 19; Length 2859;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2650; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1036 GAAACCGGTGACACCCCGGCGTCCAGAGCAGTGGCCACACAGTGGCGACGTGCCCGCC 1095
Db 199 GATGCCAGGACACCCCGGCGTCCAGAGCAGTGGCCACACAGTGGCGACGTGCCCGCC 258
QY 1096 AACAGCCGCTTCGATTGCGCCCTGACAAAGGCCATCACCCAGGAAACAGTGCAGGCGCCG 1155
Db 259 AACAGCCGCTTCGATTGCGCCCTGACAAAGGCCATCACCCAGGAAACAGTGCAGGCGCCG 318
QY 1156 GGCTGCTGTCATCCCTGCAAGAGCAGGGCTGCAAGGAGCCAGATGGGGGAGCGCTGG 1215
Db 319 GGCTGCTGTCATCCCTGCAAGAGCAGGGCTGCAAGGAGCCAGATGGGGGAGCGCTGG 378
QY 1216 TGCTTTTCCACCCAGCTTACCCAGCTACAGCTTACAGCTTGGAGAACCTGAGCTCTCTGAAATG 1275
Db 379 TGCTTTTCCACCCAGCTTACCCAGCTTACAGCTTGGAGAACCTGAGCTCTCTGAAATG 438
QY 1276 GGCTACAGCGGCACCCCTGACCCCGTACCCACCCCTTCTTCCCAAGGACATCCTGACC 1335
Db 439 GGCTACAGCGGCACCCCTGACCCCGTACCCACCCCTTCTTCCCAAGGACATCCTGACC 498
QY 1336 CTGGCGCTGGAGCTGATGATGAGAGCTGAGAACCCGCTCCACATTCACGATCAAGATCCA 1395
Db 499 CTGGCGCTGGAGCTGATGATGAGAGCTGAGAACCCGCTCCACATTCACGATCAAGATCCA 558
QY 1396 GCTAACAGCGCTACAGGTGCTTGGAGACCCCGGTGTCCACAGCGCGGACACCGTCC 1455
Db 559 GCTAACAGCGCTACAGGTGCTTGGAGACCCCGGTGTCCACAGCGCGGACACCGTCC 618
QY 1456 CCACCTCTACAGCGTGGAGTTCTCGAGGAGCCCTTTCGGGGTGATCGTGACCGGAGCTG 1515
Db 619 CCACCTCTACAGCGTGGAGTTCTCTGAGGAGCCCTTTCGGGGTGATCGTGACCGGAGCTG 678
QY 1516 GACGCGCGGTGCTGCTGAAACAGACGAGTGGCGGCCCTGTTCTTTGGGAGACAGTTCCTT 1575
Db 679 GACGCGCGGTGCTGCTGAAACAGACGAGTGGCGGCCCTGTTCTTTGGGAGACAGTTCCTT 738
QY 1576 CAGCTGTCCACTCGCTCGCTCGCAGTATATACAGGCTCGCGGAGCACCTCAGTCC 1635
Db 739 CAGCTGTCCACTCGCTCGCTCGCAGTATATACAGGCTCGCGGAGCACCTCAGTCC 798
QY 1636 CTGATGCTCAGCACCCAGCTGGACAGGATCACCTGTGGAAACCGGAGACCTTGGCGCCAG 1695
Db 799 CTGATGCTCAGCACCCAGCTGGACAGGATCACCTGTGGAAACCGGAGACCTTGGCGCCAG 858
QY 1696 CCGGTGGAGACCTTACGGGTCTCACCTTTTCTACCTGGCGCTGGAGACCGGCGGTG 1755
Db 859 CCGGTGGAGACCTTACGGGTCTCACCTTTTCTACCTGGCGCTGGAGACCGGCGGTG 918
QY 1756 GCACAGCGGGTTCCTCTGTAACAGCAATGCATGATGTGCTGCTGAGCGGAGCCCT 1815
Db 919 GCACAGCGGGTTCCTCTGTAACAGCAATGCATGATGTGCTGCTGAGCGGAGCCCT 978
QY 1816 GCGCTTAGCTGAGGTGACAGGTGGGATCTCTGATGTCTATCATCTTCTGGGCGCCAGAG 1875
Db 979 GCGCTTAGCTGAGGTGACAGGTGGGATCTCTGATGTCTATCATCTTCTGGGCGCCAGAG 1038
QY 1876 CCAAAGAGCGTGGTGCAGAGTACTGGAAGTGGGATACCCGTTGATGGCGGCATAC 1935
Db 1039 CCAAAGAGCGTGGTGCAGAGTACTGGAAGTGGGATACCCGTTGATGGCGGCATAC 1098
QY 1936 TGGGCGCTGGGCTTCCAGCTGCGCGTGGGCTACTCTCCACCGCTATCACCGCCAG 1995
Db 1099 TGGGCGCTGGGCTTCCAGCTGCGCGTGGGCTACTCTCCACCGCTATCACCGCCAG 1158


```

; PRIOR FILING DATE: 2003-06-20
;
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 2937
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-IDU fusion sequence
US-812-849-8

```

Query Match	26.4%;	Score 978.8;	DB 19;	Length 1398;
Best Local Similarity	99.3%;	Pred. No. 1.6e-236;		
Matches 983; Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0;
Qy	59	TGCTCGGATCTTACTCGCGGAGAAAGAACAGAGCCCAAGCGTCCCGGAAACGGAGTCCG	118	
Db	2	TGGGGGGTCTTACTCGCGGAGAAAGAACAGAGCCCAAGCGTCCCGGAAACGGAGTCCG	61	
Qy	119	GAGAGGAGTTCCGCATCGAGAGAGTTGAAACAGAGCTGTGGGAGAGAGGCCAGCGACTGCATC	178	
Db	62	GAGAGGAGTTCCGCATCGAGAGAGTTGAAACAGAGCTGTGGGAGAGAGGCCAGCGACTGCATC	121	
Qy	179	TTCTCTCCGTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAAC	238	
Db	122	TTCTCTCCGTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAAC	181	
Qy	239	TGCGCTTGGAGAAACTTAAAGCTTGACGGCTTGACAGAGTGGGGAGAGGAGACCGAGAC	298	
Db	182	TGCGCTTGGAGAAACTTAAAGCTTGACGGCTTGACAGAGTGGGGAGAGGAGACCGAGAC	241	
Qy	299	TCATAGCGAACCTCAATGTGTCATCTTGCGCAAGTATGGTCTGGACGGAGAAAGAGACGCTC	358	
Db	242	TCAATAGCGAACCTCAATGTGTCATCTTGCGCAAGTATGGTCTGGACGGAGAAAGAGACGCTC	301	
Qy	359	GGCAGGTGACCGACCAACTCCCTCAGTGGGACCCAGGAGACGGGCTGGATGACCCCAAGC	418	
Db	302	GGCAGGTGACCGACCAACTCCCTCAGTGGGACCCAGGAGACGGGCTGGATGACCCCAAGC	361	
Qy	419	TGAAAAAGCTGTGGCACAGGCGGAGACCTCTGGGAAATTCCTCGGCGAAGAACTGGACA	478	
Db	362	TGAAAAAGCTGTGGCACAGGCGGAGACCTCTGGGAAATTCCTCGGCGAAGAACTGGACA	421	
Qy	479	AGCTCTGGCGGGAGTTCCTTGCATCACAAAGAGAAAGTTACACGAGTACCAACGTCCTGCTGG	538	
Db	422	AGCTCTGGCGGGAGTTCCTTGCATCACAAAGAGAAAGTTACACGAGTACCAACGTCCTGCTGG	481	
Qy	539	AGACCCCTGAGCAGGACCGGAAGAAATCCAGAGAACGTCATTAGCCCTCGGACCTGAGCG	598	

Db 482 AGACCTTGAGCAGGACCGAAGAAATCCACGAGAACGTCTATTAGCCCTCGGACCTGAGCG 541
Qy 599 ACATCAAGGCGACGCTCTGCACAGCAGGACACACGAGCTGGAAGGAGAGCTGCGCAGCA 658
Db 542 ACATCAAGGCGACGCTCTGCACAGCAGGACACACGAGCTGGAAGGAGAGCTGCGCAGCA 601
Qy 659 TCAACAGGCGCTGGAACCGCTCGCAGGCTGAGCAGCAGGCTACAGCAGCTGAGGCTG 718
Db 602 TCAACAGGCGCTGGAACCGCTCGCAGGCTGAGCAGCAGGCTACAGCAGCTGAGGCTG 661
Qy 719 AGTTGAGGAGCCAGGCTGATTGACCTGTGGACCTGCGCAGCTGCGCAACCTCAGCG 778
Db 662 AGTTGAGGAGCCAGGCTGATTGACCTGTGGACCTGCGCAGCTGCGCAACCTCAGCG 721
Qy 779 ACAAGAGCTGGAAGCGCTTCCGGGAGGAGCTCAAGCAGCTTCAAGCAACCAAAATCGAGAAGC 838
Db 722 ACNAGAGCTGGAAGCGCTTCCGGGAGGAGCTCAAGCAGCTTCAAGCAACCAAAATCGAGAAGC 781
Qy 839 ACAACCACTACCAAGAGCAGCTGGAGATTGCGCACGAGAACTGAGGCAACGAGAGCG 898
Db 782 ACAACCACTACCAAGAGCAGCTGGAGATTGCGCACGAGAACTGAGGCAACGAGAGCG 841
Qy 899 TGGGCGACCGGAGCTGTGAGCGCGAGCGCGAGAGCAACGCGCTGCTGGAGGGCGGA 958
Db 842 TGGGCGACCGGAGCTGTGAGCGCGAGCGCGAGAGCAACGCGCTGCTGGAGGGCGGA 901
Qy 959 CCAAGAGCTGGGCTACAGCTGAAGAGCAGCTGCGAGCAGCTGCGGAGGAGTCTCCA 1018
Db 902 CCNAGAGCTGGGCTACAGCTGAAGAGCAGCTGCGAGCAGCTGCGGAGGAGTCTCCA 961
Qy 1019 GAGTCGCGCCGAGGAGGAGAAACCGGTGCAC 1048
Db 962 GAGTCGCGCCGAGGAGGAGAAACCGGTTCAC 991

RESULT 10

US-10-812-849-10
; Sequence 10, Application US/10812849
; Publication No. US2005004227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND
; FILE OF INVENTION: OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; CURRENT FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GDNF fusion sequence
US-10-812-849-10

Query Match 26.4%; Score 978.8; DB 19; Length 1398;

Best Local Similarity 99.3%; Pred. No. 1.6e-236;

Matches 983; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 59 TGTCTGGATCCTACTCGCGGAGAGAACCCAGCCCGCTCCCGAAACCGGAGTCCG 118
Db 2 TGGGGGGTCTTACTCGCGGAGAGAACCCAGCCCGCTCCCGAAACCGGAGTCCG 61
Qy 119 GAGAGAGTTCCTGAGAGAGTTCGAACCAAGCTGTGGGAGAGGCGCCAGCAGTGCATC 178
Db 62 GAGAGAGTTCCTGAGAGAGTTCGAACCAAGCTGTGGGAGAGGCGCCAGCAGTGCATC 121
Qy 179 TTCTCTCCGCTGAGGCTGGCGAGCTCCACGCTGATCTGAAGATACAGGAGAGGAGCAAC 238

Db 122 TTCCTCCCTGAGGCTGCGCGAGCTCCACGCTGATCTGAAGATACAGGAGAGGAGCAAC 181
Qy 239 TCGCTTGGAGAGAACTAAAGCTTTGACGCTTTGACGAGATGGGAGAGAGGAGCGAGAC 298
Db 182 TCGCTTGGAGAGAACTAAAGCTTTGACGCTTTGACGAGATGGGAGAGAGGAGCGAGAC 241
Qy 299 TCATACGCAACCTCAATGTCATCTTGGCCAAAGTATGGTCTGACGCGAAAGAGGAGCGCTC 358
Db 242 TCATACGCAACCTCAATGTCATCTTGGCCAAAGTATGGTCTGACGCGAAAGAGGAGCGCTC 301
Qy 359 GCGAGTGCACGAGCAACCTCCCTCAGTGGCCACCCAGGAGACCGGCTGGATGACCCGAGGC 418
Db 302 GCGAGTGCACGAGCAACCTCCCTCAGTGGCCACCCAGGAGACCGGCTGGATGACCCGAGGC 361
Qy 419 TGGAAAGCTGTGGCACAAGGCGCAAGCTCTGGGAAATTTCTCCGCGCAAGAACTGGAACA 478
Db 362 TGGAAAGCTGTGGCACAAGGCGCAAGCTCTGGGAAATTTCTCCGCGCAAGAACTGGAACA 421
Qy 479 AGCTCTGGGGGAGTTCCTGCAATCAAAAGAGAAAGTTTCCAGAGTACAACTGCTCTGCTG 538
Db 422 AGCTCTGGGGGAGTTCCTGCAATCAAAAGAGAAAGTTTCCAGAGTACAACTGCTCTGCTG 481
Qy 539 AGACCTCTGAGCAGGACCGGAGAAATCCACGAGAACGTCTATTAGCCCTCGGACCTGAGCG 598
Db 482 AGACCTCTGAGCAGGACCGGAGAAATCCACGAGAACGTCTATTAGCCCTCGGACCTGAGCG 541
Qy 599 ACATCAAGGCGAGCTCTGTCAGCAGCAGGACACACGAGCTGAAGGAGAGAGTTCGCGAGCA 658
Db 542 ACATCAAGGCGAGCTCTGTCAGCAGCAGGACACACGAGCTGAAGGAGAGAGTTCGCGAGCA 601
Qy 659 TCAACAGGCGCTGAGCAGCGCTGCGCAGGCTGAGCCACAGGCGCTACAGCAGTGAAGGCTG 718
Db 602 TCAACAGGCGCTGAGCAGCGCTGCGCAGGCTGAGCCACAGGCGCTACAGCAGTGAAGGCTG 661
Qy 719 AGTTGAGGAGCCAGGCTGATTGACCTGTGGGAGCTTGGGAGCTCCGCCAACCTCAGCG 778
Db 662 AGTTGAGGAGCCAGGCTGATTGACCTGTGGGAGCTTGGGAGCTCCGCCAACCTCAGCG 721
Qy 779 ACAAGAGCTGAGGCGCTTCCGGGAGGAGCTCAAGCAGCTTCAAGCAACCAAAATCGAGAAGC 838
Db 722 ACAAGAGCTGAGGCGCTTCCGGGAGGAGCTCAAGCAGCTTCAAGCAACCAAAATCGAGAAGC 781
Qy 839 ACAACCACTACCAAGAGCAGCTGGAAGTTGCGCACGAGAACTGAGGCAACGAGAGCG 898
Db 782 ACAACCACTACCAAGAGCAGCTGGAAGTTGCGCACGAGAACTGAGGCAACGAGAGCG 841
Qy 899 TGGGCGACCGGAGCGCTGTGAGCGCGAGGAGAGCAACGCGCTGCTGGAGGGCGGA 958
Db 842 TGGGCGACCGGAGCGCTGTGAGCGCGAGGAGAGCAACGCGCTGCTGGAGGGCGGA 901
Qy 959 CCAAGGAGCTGGGCTACAGCTGAAGAGCAGCTTGCAGGAGCTGTCGCGGAGGAGTCTCCA 1018
Db 902 CCNAGAGCTGGGCTACAGCTGAAGAGCAGCTGCGAGGAGCTGTCAGGAGGAGTCTCCA 961
Qy 1019 GAGTCGCGCCGAGGAGGAGAAACCGGTGCAC 1048
Db 962 GAGTCGCGCCGAGGAGGAGAAACCGGTTCAC 991

RESULT 11

US-10-172-118-757

; Sequence 757, Application US/10172118

; Publication No. US20030224374A1

; GENERAL INFORMATION:

; APPLICANT: Dai, Hongyue

; APPLICANT: He, Yudong

; APPLICANT: Lineley, Peter

; APPLICANT: Mao, Mao

; APPLICANT: Roberts, Chris

; APPLICANT: Van 't Veer, Laura

; APPLICANT: Van de Vijver, Marc

; APPLICANT: Bernards, Rene

; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

	Query Match	25.94;	Score 957;	DB 17;	Length 1493;
	Best Local Similarity	99.0%;	Pred. No. 5.2e-231;		
	Matches 963;	Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0;
469	AAAGCTGTGGCAACAAGGCGAAGACCTCTGGGAAATCTTCGGGCGAAGAACTGGACAAAGCT	528			
483	CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTTCACGAGTACAAAGCTCTCTCTGGAGAC	542			
529	CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTTCACGAGTACAAAGCTCTCTCTGGAGAC	588			
543	CTTGACGAGGACCGAAGAAATCCACGAGAACGTCATTTAGCCCTCTCGGACCTTGAGCGACAT	602			
589	CTTGACGAGGACCGAAGAAATCCACGAGAACGTCATTTAGCCCTCTCGGACCTTGAGCGACAT	648			
603	CAAGGCGACGCTCTTGCAACAGACAGCAACACGAGCTGAAGGAGAAAGTTGGCGACATCAA	662			
649	CAAGGCGACGCTCTTGCAACAGACAGCAACACGAGCTGAAGGAGAAAGTTGGCGACATCAA	708			
663	CAAGGCGCTGGACCGCTCTGGCAGAGGCTCAGCCACAGGCGCTACAGCACTGAGGCTGAGTT	722			
709	CAAGGCGCTGGACCGCTCTGGCAGAGGCTCAGCCACAGGCGCTACAGCACTGAGGCTGAGTT	768			
723	CGAGGAGCCACGGGTGATTGACCTGTGGGACCTGGCGCAGTCCGCCACCTCAGCGACAA	782			
769	CGAGGAGCCACGGGTGATTGACCTGTGGGACCTGGCGCAGTCCGCCACCTCAGCGACAA	828			
783	GGAGCTTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAAATCGAGAAGACAA	842			
829	GGAGCTTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAAATCGAGAAGACAA	888			
843	CCACTACCAAGACGAGCTGGAGATTGGCCACGAGAACGTGAGGCAACGCGAGAGCGTGGG	902			

QY 423 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAATTTCTCGGCGAAGAACTGGACAAGCT 482
Db 469 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAATTTCTCGGCGAAGAACTGGACAAGCT 528
QY 483 CTGGCGGGAGTTCTCTGCATCACAAGAGAGAAAGTTTACAGAGTCAACAGTCTCTGTGGAGAC 542
Db 529 CTGGCGGGAGTTCTCTGCATCACAAGAGAGAAAGTTTACAGAGTCAACAGTCTCTGTGGAGAC 588
QY 543 CTGTAGCAGGACCGAAGAAATCCAGAGAAAGTTCATTAGCCCTTCGGACCTTGAGCGACAT 602
Db 589 CTTGAGCAGGACCGAAGAAATCCAGAGAAAGTTCATTAGCCCTTCGGACCTTGAGCGACAT 648
QY 603 CTAAGGCGCGTCTCTGCACAGCAGGACACACGGAGCTGAGAGAAAGTTCGGCAGCATCAA 662
Db 649 CTAAGGCGCGTCTCTGCACAGCAGGACACACGGAGCTGAGAGAAAGTTCGGCAGCATCAA 708
QY 663 CTAAGGCGCGTCTCTGCACAGCAGGAGTTCAGCCACAGGAGTTCAGACACTGAGGCTGAGTT 722
Db 709 CTAAGGCGCGTCTCTGCACAGGAGTTCAGCCACAGGAGTTCAGACACTGAGGCTGAGTT 768
QY 723 CGAGGAGCCAGGAGTTCAGCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCT 782
Db 769 CGAGGAGCCAGGAGTTCAGCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCT 828
QY 783 GGAGCTGGAGGCGTTCCTGGGAGGAGCTCAAGCACTTCGAAGCCAAATTCGAGAGCACAA 842
Db 829 GGAGCTGGAGGCGTTCCTGGGAGGAGCTCAAGCACTTCGAAGCCAAATTCGAGAGCACAA 888
QY 843 CCACTACCAAGAGCAGCTGGAGATTTCGCACAGAGAGCTGAGGACGCGAGAGCGTGGG 902
Db 889 CCACTACCAAGAGCAGCTGGAGATTTCGCACAGAGAGCTGAGGACGCGAGAGCGTGGG 948
QY 903 CGACGCGGAGCGTGTGAGCGGACGCGGAGAGAGCTGAGGACGCGGAGAGCGTGGG 962
Db 949 CGACGCGGAGCGTGTGAGCGGACGCGGAGAGAGCTGAGGACGCGGAGAGCGTGGG 1008
QY 963 GGAGCTGGGCTACACGCTGAAGAGCATCTGAGGAGCTGTCCGGCAGAGTCTCCAGAGC 1022
Db 1009 GGAGCTGGGCTACACGCTGAAGAGCATCTGAGGAGCTGTCCGGCAGAGTCTCCAGAGC 1068
QY 1023 TCGCGCGGAGGCA 1035
Db 1069 TCGGCACAACGAA 1081

RESULT 13

US-10-342-887-757
; Sequence 757, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 757
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-342-887-757

Query Match 25.9%; Score 957; DB 17; Length 1493;
Best Local Similarity 99.0%; Pred. No. 5.2e-231;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 63 CGGATCCTACTCGCGGAGAGAACAGCCAGCGTCCCGAAGCGGAGTCCGGAGA 122
Db 109 CGGCAAGTACTCGCGGAGAGAAACAGCCCAAGCGTCCCGAAGCGGAGTCCGGAGA 168
QY 123 GGAGTTTCCGATCGGAGAGTTTGAACAGCTGTGGGAGAGGCGCCAGCTGCTATCTCC 182
Db 169 GGAGTTTCCGATCGGAGAGTTTGAACAGCTGTGGGAGAGGCGCCAGCTGCTATCTCC 228
QY 183 TCCGTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAATCGC 242
Db 229 TCCGTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAATCGC 288
QY 243 CTGGAGAGAACTAAAGCTTGACGGCTTGACCAAGATGGGAGAGGAGCGAGCTCAT 302
Db 289 CTGGAGAGAACTAAAGCTTGACGGCTTGACCAAGATGGGAGAGGAGCGAGCTCAT 348
QY 303 ACGCAACCTCAATGTCTCTTGGCCAAAGTATGGTCTTGAGACGGAAGAGAGCGCTCGGCA 362
Db 349 ACGCAACCTCAATGTCTCTTGGCCAAAGTATGGTCTTGAGACGGAAGAGAGCGCTCGGCA 408
QY 363 GGTGACCAAGCAACTCCCTCAGTGGCACCCAGAGAGCGGGCTGGATGACCCAGGCTGGA 422
Db 409 GGTGACCAAGCAACTCCCTCAGTGGCACCCAGAGAGCGGGCTGGATGACCCAGGCTGGA 468
QY 423 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAAAATTCCTCGGCGAAGAGTTCGACAGCT 482
Db 469 AAAGCTGTGGCACAAGCGGAGAGCTCTGGGAAAATTCCTCGGCGAAGAGTTCGACAGCT 528
QY 483 CTGGCGGAGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAAGAGTACAAAGTCTCTGTGGAGAC 542
Db 529 CTGGCGGAGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAAGAGTACAAAGTCTCTGTGGAGAC 588
QY 543 CTTGAGCAGGACCGAAGAAATCCAGAGAACTTCATTAGCCCTTCGGACCTTGAGCGACAT 602
Db 589 CTTGAGCAGGACCGAAGAAATCCAGAGAACTTCATTAGCCCTTCGGACCTTGAGCGACAT 648
QY 603 CTAAGGCGCGTCTCTGCACAGCAGGACACCGGAGCTGAGAGAGAGTTCGGCAGCATCAA 662
Db 649 CTAAGGCGCGTCTCTGCACAGCAGGACACCGGAGCTGAGAGAGAGTTCGGCAGCATCAA 708
QY 663 CTAAGGCGCGTCTCTGCACAGGAGTTCAGCCACAGGAGTTCAGACACTGAGGCTGAGTT 722
Db 709 CTAAGGCGCGTCTCTGCACAGGAGTTCAGCCACAGGAGTTCAGACACTGAGGCTGAGTT 768
QY 723 CGAGGAGCCAGGAGTTCAGCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCT 782
Db 769 CGAGGAGCCAGGAGTTCAGCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCTGTGGGACCT 828
QY 783 GGAGCTGGAGGCGTTCCTGGGAGGAGCTCAAGCACTTCGAAGCCAAATTCGAGAGCACAA 842
Db 829 GGAGCTGGAGGCGTTCCTGGGAGGAGCTCAAGCACTTCGAAGCCAAATTCGAGAGCACAA 888
QY 843 CCACTACCAAGAGCAGCTGGAGATTTCGCACAGAGAGCTGAGGACGCGAGAGCGTGGG 902
Db 889 CCACTACCAAGAGCAGCTGGAGATTTCGCACAGAGAGCTGAGGACGCGAGAGCGTGGG 948
QY 903 CGACGCGGAGCGTGTGAGCGGACGCGGAGAGAGCTGAGGACGCGGAGAGCGTGGG 962
Db 949 CGACGCGGAGCGTGTGAGCGGACGCGGAGAGAGCTGAGGACGCGGAGAGCGTGGG 1008
QY 963 GGAGCTGGGCTACACGCTGAAGAGCATCTGAGGAGCTGTCCGGCAGAGTCTCCAGAGC 1022
Db 1009 GGAGCTGGGCTACACGCTGAAGAGCATCTGAGGAGCTGTCCGGCAGAGTCTCCAGAGC 1068
QY 1023 TCGCGCGGAGGCA 1035
Db 1069 TCGGCACAACGAA 1081

```
RESULT 14
US-10-472-516-3
; Sequence 3, Application US/10472516
; Publication No. US20040248785A1
; GENERAL INFORMATION:
; APPLICANT: Saenko, Evgeni L.
; APPLICANT: Sarafanov, Andrey G.
; TITLE OF INVENTION: Methods and Compositions for Reducing Heparan Sulfate
; TITLE OF INVENTION: Proteoglycan-Mediated Clearance of Factor VIII
; FILE REFERENCE: 1327.066001
; CURRENT APPLICATION NUMBER: US/10/472,516
; PRIOR FILING DATE: 2003-09-24
; PRIOR APPLICATION NUMBER: PCT/US02/00583
; PRIOR FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 60/260,904
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 3
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (14)..(115)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (116)..(1084)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14)..(1084)
US-10-472-516-3

Query Match 25.9%; Score 957; DB 18; Length 1493;
Best Local Similarity 99.0%; Pred. NO. 5.2e-231;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTACTCGCGGAGAGAACACAGCCCGTCCCGAAACGGAGTCCGGAGA 122
DB 109 CGGAAAGTACTCGCGGAGAGAACACAGCCCGTCCCGAAACGGAGTCCGGAGA 168
QY 123 GGAGTTCCGATGGAGAAGTTGAACAGCTGTGGGAGAGGCCGCGACTGCATCTCC 182
DB 169 GGAGTTCCGATGGAGAAGTTGAACAGCTGTGGGAGAGGCCGCGACTGCATCTCC 228
QY 183 TCCGCTGAGGCTGGCCGAGCTCAGCGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 242
DB 229 TCCGCTGAGGCTGGCCGAGCTCAGCGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 288
QY 243 CTGGAGAGAACTAAAGCTTACGGCTTGGACGAAGATGGGAGAGGAGGAGCAACTCAT 302
DB 289 CTGGAGAGAACTAAAGCTTACGGCTTGGACGAAGATGGGAGAGGAGGAGCAACTCAT 348
QY 303 ACGCAACCTCAATGTCTATCTTGGCCCAAGTATGTCTGGACGGAAGAGGAGCGCTCGGCA 362
DB 349 ACGCAACCTCAATGTCTATCTTGGCCCAAGTATGTCTGGACGGAAGAGGAGCGCTCGGCA 408
QY 363 GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGACGGGCTGGATGATCCCGAGGCTGGA 422
DB 409 GGTGACCAAGCACTCCCTCAGTGGGACCCAGGAGACGGGCTGGATGATCCCGAGGCTGGA 468
QY 423 AAAGCTGTGGCAACAAGCGAAGCACTCTGGGAAATTTCTCCGGCGAAGCACTGCAAGCT 482
DB 469 AAAGCTGTGGCAACAAGCGAAGCACTCTGGGAAATTTCTCCGGCGAAGCACTGCAAGCT 528
QY 483 CTGGCGGAGGTTCTCTGCATCACAAGAGAGAAAGTTACAGAGTACAACTGCTCTGGAGAC 542
DB 529 CTGGCGGAGGTTCTCTGCATCACAAGAGAGAAAGTTACAGAGTACAACTGCTCTGGAGAC 588
QY 543 CTTGAGCAGGACCGAAGAAATCCACGAGAAAGCTCATTAGCCCTCGGACCTGAGCGACAT 602
DB 543 CTTGAGCAGGACCGAAGAAATCCACGAGAAAGCTCATTAGCCCTCGGACCTGAGCGACAT 602

RESULT 15
US-10-741-600-440
; Sequence 440, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 440
; LENGTH: 2064
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-440

Query Match 25.8%; Score 954.2; DB 19; Length 2064;
Best Local Similarity 98.3%; Pred. NO. 2.7e-230;
Matches 956; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTACTCGCGGAGAGAGAACACCGCCCAAGCCGTCCCGAAACGGAGTCCGGAGA 122
DB 146 CGGCAAGTACTCGCGGAGAGAGAACACCGCCCAAGCCGTCCCGAAACGGAGTCCGGAGA 205
QY 123 GGAGTTCCGCATGGAGAAAGTTGAACAGCTGTGGGAGAGGCCCGCACTGCATCTTCC 182
DB 206 GGAGTTCCGCATGGAGAAAGTTGAACAGCTGTGGGAGAGGCCCGCACTGCATCTTCC 265
QY 183 TCCCGTGAAGTGGCCGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 242
DB 266 TCCCGTGAAGTGGCCGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCAACTCGC 325
QY 243 CTGGAGAGAACTAAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 302
DB 326 CTGGAGAGAACTAAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 385
```


QY	303	ACGCAACCTCAATGTTCATCTTGCCCAAGTATGCTTGGAAGGAAAGAAAGGACGCTCGGCA	362
Db	386	ACGCARCTCARTGTTCATCTTGCCCAAGTATGCTTGGAAGGAAAGAAAGGACGCTCGGCA	445
QY	363	GGTGACGACGACTCCCTCAGTGGCACCACGGAAGAGGGCTGGATGACCCGAGGCTGGA	422
Db	446	GGTGACGACGACTCCCTCAGTGGCACCACGGAAGAGGGCTGGATGACCCGAGGCTGGA	505
QY	423	AAAGCTGTGGCACAAGCGAAGACCTCTGGGAAATTCCTCGGCGAAAGAACTGGACAAGCT	482
Db	506	AAAGCTGTGGCACAAGCGGAGACCTCTGGGAAATTCCTCGGCGAAGAACTGGACAAGCT	565
QY	483	CTGGCGGGAGTTCTTGATCACAAGAGAAAGTTCAAGTACAAAGTCTCTGCTGGAGAC	542
Db	566	CTGGCGGGAGTTCTTGATCACAAGAGAAAGTTCAAGTACAAAGTCTCTGCTGGAGAC	625
QY	543	CTTGACGACGACCGAGAGAAATCCACGAGACGTCATTAGCCCTCGGACCTGAGCGACAT	602
Db	626	CTTGACGACGACCGAGAGAAATCCACGAGACGTCATTAGCCCTCGGACCTGAGCGACAT	685
QY	603	CAAGGGCAGCGTCTCTGCACAGCAGGACACGAGAGCTGAAGGAGAACTGCGCAGCATCAA	662
Db	686	CAAGGGCAGCGTCTCTGCACAGCAGGACACGAGAGCTGAAGGAGAACTGCGCAGCATCAA	745
QY	663	CCAGGGCTGGACCCGCTCGCAGGGTCAAGCCACGAGGGCTACAGCACTGAGGCTGAGTT	722
Db	746	CCAGGGCTGGACCCGCTCGCAGGGTCAAGCCACGAGGGCTACAGCACTGAGGCTGAGTT	805
QY	723	CGAGGAGCCGAGGGTGTGACCTGTGGACCTGGGCGCAGTCCGCCAACCTCAGCGACAA	782
Db	806	CGAGGAGCCGAGGGTGTGACCTGTGGACCTGGGCGCAGTCCGCCAACCTCAGCGACAA	865
QY	783	GGAGCTGGAGGGCTTCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCACAA	842
Db	866	GGAGCTGGAGGGCTTCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCACAA	925
QY	843	CCACTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGCACGCAAGAGCGTGGG	902
Db	926	CCACTACCAAGACAGCTGGAGATTGCGCACGAGAAAGCTGAGGCACGCAAGAGCGTGGG	985
QY	903	CGACGGCGAGCGTGTGAGCCGCGCAGCGGAGAGCAAGCCCTGCTGGAGGGCGGACCAA	962
Db	986	CGAYGGCGAGCGTGTGAGCCGCGCAGCGGAGAGCAAGCCCTGCTGGAGGGCGGACCAA	1045
QY	963	GGAGCTGGGCTACACGCTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC	1022
Db	1046	GGAGCTGGGCTACACGCTGAAGAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC	1105
QY	1023	TCGCGCCGAGGCA	1035
Db	1106	TCGGCACACGAA	1118

Search completed: April 5, 2005, 00:59:58
Job time : 7246 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2005, 11:07:54 ; Search time 10485 Seconds
(without alignments)
13439.573 Million cell updates/sec

Title: US-10-600-862A-6
Perfect score: 3702
Sequence: 1 ctaccgcacgtcggggtcc.....tcagtggtgtgactcgag 3702

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gss1:
9: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2526.6	68.2	2859	9 AY413853	AY413853 Homo sapi
2	1898.2	51.3	3488	3 AK088481	AK088481 Mus muscu
3	1898.2	51.3	3493	3 AK052211	AK052211 Mus muscu
4	1881.6	50.8	2859	9 AY413854	AY413854 Pan trogl
5	1802.6	48.7	2862	9 AY413855	AY413855 Mus muscu
6	959.8	25.9	1100	5 BM922403	BM922403 AGENCOURT
7	957	25.9	1347	3 CR617514	CR617514 full-leng
8	957	25.9	1373	3 CR603866	CR603866 full-leng
9	957	25.9	1385	3 CR619388	CR619388 full-leng
10	957	25.9	1392	3 CR606004	CR606004 full-leng
11	957	25.9	1399	3 CR596948	CR596948 full-leng
12	957	25.9	1399	3 CR625424	CR625424 full-leng
13	957	25.9	1402	3 CR591537	CR591537 full-leng
14	957	25.9	1409	3 CR619887	CR619887 full-leng
15	957	25.9	1411	3 CR607197	CR607197 full-leng
16	957	25.9	1418	3 CR616455	CR616455 full-leng
17	957	25.9	1427	3 CR618610	CR618610 full-leng
18	957	25.9	1430	3 CR607312	CR607312 full-leng
19	957	25.9	1439	3 CR624552	CR624552 full-leng
20	957	25.9	1442	3 CR625692	CR625692 full-leng
21	957	25.9	1444	3 CR605976	CR605976 full-leng
22	957	25.9	1449	3 CR597384	CR597384 full-leng
23	957	25.9	1449	3 CR601788	CR601788 full-leng
24	957	25.9	1449	3 CR611476	CR611476 full-leng

25	957	25.9	1451	3	CR604735	CR604735 full-leng
26	957	25.9	1451	3	CR625232	CR625232 full-leng
27	957	25.9	1454	3	CR603821	CR603821 full-leng
28	957	25.9	1456	3	CR591442	CR591442 full-leng
29	957	25.9	1458	3	CR599568	CR599568 full-leng
30	957	25.9	1460	3	CR604363	CR604363 full-leng
31	957	25.9	1460	3	CR614498	CR614498 full-leng
32	957	25.9	1463	3	CR626102	CR626102 full-leng
33	957	25.9	1464	3	CR607263	CR607263 full-leng
34	957	25.9	1465	3	CR613053	CR613053 full-leng
35	927.2	25.0	1453	3	CR617840	CR617840 full-leng
36	899.2	24.3	968	1	AL551402	AL551402 full-leng
37	894.4	24.2	1004	5	BX337935	BX337935 full-leng
38	884.6	23.9	1011	5	BX458886	BX458886 full-leng
39	883	23.9	961	5	BX379126	BX379126 full-leng
40	882	23.8	1003	1	AL518587	AL518587 full-leng
41	880	23.8	1020	1	AL542049	AL542049 full-leng
42	877.4	23.7	1013	1	AL518778	AL518778 full-leng
43	877	23.7	1023	5	BX399454	BX399454 full-leng
44	873.6	23.6	991	5	BX381548	BX381548 full-leng
45	867.4	23.4	990	1	AL550423	AL550423 full-leng

ALIGNMENTS

RESULT 1
AY413853
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
ORIGIN

AY413853 2859 bp DNA linear GSS 17-DEC-2003
Homo sapiens GAA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
AY413853.1 GI:39769815
GSS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2859)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarawal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1..2859
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2859
/gene="GAA"
/locus_tag="HCM5029"

Query Match 68.2%; Score 2526.6; DB 9; Length 2859;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 2532; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 1036 GAAACGGTGCACACCCCGCGTCCCGAGAGTCCACAGTGCACGTCCCGCC 1095
|||||


```

QY 3256 CTGCGGCGCCCTGTGACACCATCAAGTCCACCTCCGGGCTGGGTACATCATCCCCCTG 3315
Db 2419 CTGCGGCGCCCTGTGACACCATCAAGTCCACCTCCGGGCTGGGTACATCATCCCCCTG 2478
QY 3316 CAGGGCCCTGGCTTCACACACAGAGTCCCGCAGCAGCCCATGCGCTGGCTGGCC 3375
Db 2479 CAGGGCCCTGGCTTCACACACAGAGTCCCGCAGCAGCCCATGCGCTGGCTGGCC 2538
QY 3376 CTAACCAAGGTGGAGAGGCCCGGAGGGAGCTTCTTGGGACGATGGAGAGCCTGGAA 3435
Db 2539 CTGACCAAGGTGGAGAGGCCCGGAGGGAGCTTCTTGGGACGATGGAGAGCCTGGAA 2598
QY 3436 GTGCTGGAGAGGGGCTTACACACAGTCACTTCTGCGCCAGGAATACACGATCGTG 3495
Db 2599 GTGCTGGAGAGGGGCTTACACACAGTCACTTCTGCGCCAGGAATACACGATCGTG 2658
QY 3496 AATGAGCTGTGTGACCTGAGGAGCTGGGCTGCAGCTGCAGAGGTGACTGTC 3555
Db 2659 AATGAGCTGTGTGACCTGAGGAGCTGGGCTGCAGCTGCAGAGGTGACTGTC 2718
QY 3556 CTGGGCGTGGCCACCGCGCCCGCAGCAGGTCTCTTCCAAAGGTGTCCTCTCCAACTTC 3615
Db 2719 CTGGGCGTGGCCACCGCGCCCGCAGCAGGTCTCTTCCAAAGGTGTCCTCTCCAACTTC 2778
QY 3616 ACTACAGCCCGACACCAAGTCTCGGACATCTGTCTCGCTGTGATGGAGAGCAG 3675
Db 2779 ACTACAGCCCGACACCAAGTCTCGGACATCTGTCTCGCTGTGATGGAGAGCAG 2838
QY 3676 TTCTCGTCAGCTGGTGT 3694
Db 2839 TTCTCGTCAGCTGGTGT 2857

RESULT 2
LOCUS AK088481
DEFINITION Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430018M07 product:glucosidase, alpha, acid, full insert sequence.
ACCESSION AK088481.1 GI:26353503
VERSION AK088481
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
AUTHORS Carninci, P., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsui, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

```

```

REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
source
1. 3488
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM DB:E430018M07"
/db_xref="taxon:10090"
/clone="E430018M07"
/cell_type="thymic cells"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stages="2 days neonate"
237. 3098
/notes="unnamed protein product; glucosidase, alpha, acid (MGD|GI:95609, GB|NM_008064, evidence: BLASTN, 99%, match=3299) putative"
/codon_start=1
/protein_id="BAC40382.1"
/db_xref="GI:26353504"
/translation="MNIRKPLCSNVGACTLISLTAVILGHMLRELMLPQDLHE SSSGUNKYTRHQBQYKGPLHIQEQTEQKPAFTQCDVPSPSPDCAPDGIGSEQ CEARGCVPAQVLKPEIQGQFPFSPSPYSLNLSSTESGTLTITSPITFF PRDVLTLQVLMETSRHLKIDPKAPSKRYEVLPTPLRVLSQAPSLYSVSEPF GIVVRKLGKRVLLNTTVAIFFAFDQFLQSLTSLPSQHTGLGHLSPALSTDWARI TLWNRDTPSQGNTLYGSHPFVLALEDGLAGHVFLLNSAMDVILQSPALSTWSTG GILDYVFLGPEPKSVWQYLDVQVYPMPPYMGGLGFLCRWGYSTALVROVENMT RTHFPDQVMDLIDINDARRDFTFNQDPSFADPPDWVRHLHQGRYRMIVDPAISAG PAGESYRYPYDEGLRGVFTITNETGQLIGKVMVPGTTAFDFTTNPTLDMQDMVSEPHA

```

```

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3488)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
Location/Qualifiers
1. 3488
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM DB:E430018M07"
/db_xref="taxon:10090"
/clone="E430018M07"
/cell_type="thymic cells"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stages="2 days neonate"
237. 3098
/notes="unnamed protein product; glucosidase, alpha, acid (MGD|GI:95609, GB|NM_008064, evidence: BLASTN, 99%, match=3299) putative"
/codon_start=1
/protein_id="BAC40382.1"
/db_xref="GI:26353504"
/translation="MNIRKPLCSNVGACTLISLTAVILGHMLRELMLPQDLHE SSSGUNKYTRHQBQYKGPLHIQEQTEQKPAFTQCDVPSPSPDCAPDGIGSEQ CEARGCVPAQVLKPEIQGQFPFSPSPYSLNLSSTESGTLTITSPITFF PRDVLTLQVLMETSRHLKIDPKAPSKRYEVLPTPLRVLSQAPSLYSVSEPF GIVVRKLGKRVLLNTTVAIFFAFDQFLQSLTSLPSQHTGLGHLSPALSTDWARI TLWNRDTPSQGNTLYGSHPFVLALEDGLAGHVFLLNSAMDVILQSPALSTWSTG GILDYVFLGPEPKSVWQYLDVQVYPMPPYMGGLGFLCRWGYSTALVROVENMT RTHFPDQVMDLIDINDARRDFTFNQDPSFADPPDWVRHLHQGRYRMIVDPAISAG PAGESYRYPYDEGLRGVFTITNETGQLIGKVMVPGTTAFDFTTNPTLDMQDMVSEPHA

```

QVFDGMWLDMEPSNFRVSGQCPNNLENPPYVVPVGVVGILOAATCASHQFLS
THYNLNLVLEATASSALVKTRTREPFIISRFSFGHRYAGYHWTGDVRSSEHL
AYSPDILQNLGLVPLVADICGFIGDTSBELCVRWTOLGAFYPMRNDHNSVPO
EVPRESETAQOAKRKAFLRYALLPYLYTLFHRHVRGDTVARPLFLPEPDPSTWSV
DRQLLWGPALLITPVLPEPKTEVGYFPKGTWYNMQMVSVDLSGLTLPSPSSASSRSDA
VQSGMWLTLEAPLDITINHLREGYIIPQGSFLSTTSRKKQPMALAVALTASGBADG
ELFWDGSLVPLERGAYLTVLFSKNNITVNLKRVRTKEGAELQRLREVLVGTATAP
TOVLSNGIPVSNFTYSPDNKSLAIPVSLMLMBELFQISWS"
3464..3469

polya_signal

/note="putative"

polya_site

/note="putative"

ORIGIN

Query Match 51.3%; Score 1898.2; DB 3; Length 3488;
Best Local Similarity 82.6%; Pred. No. 0;
Matches 2186; Conservative 0; Mismatches 458; Indels 3; Gaps 1;
1048 CACCCCGCGCTCCAGAGCAGTGCACACACAGTGCAGCTCCCCCCCCAACAGCGCTTC 1107
1108 GATTGCGCCCTGACAGGCCATCACCCAGGAACAGTGCAGGCCCGCGGCTGCTGTAC 1167
507 GACTGTGCCCCACAAAGGCATCTCACAGGAGCAATGCAGGCCCGCGGCTGCTGTAT 566
1168 ATCCCTGCAAGCAGGGCTGCAGGAGGCCAGATGGGCGAGCCCTGTGTCTTCCCA 1227
567 GTCCAGCAGGCGAGTGTGAAGGAGCCGAGATAGGCGAGCCCTGTGTCTTCCCT 626
1228 CCAGCTACCCAGCTACAAAGTGGAGAACCTGAGCTCTCTGAAATGGGCTACAGGCC 1287
627 CCAGCTACCAAGCTACCGTCTAGAGAACCTGAGCTCTACAGAGTCGGGGTACACAGCC 686
1288 ACCCTACCGGTACACCCCGACCTTCTCCCAAGGACATCCTGACCTCGGCTGGAC 1347
687 ACCCTACCGGTACACCCCGACCTTCTCCCAAGGATGTGTGACCTTACAGCTGGAG 746
1348 GTGATGATGGAGACTGAGAACCGCTTCCACTTCAAGATCAAGATCCAGCTAACAGGCG 1407
747 GTGCTGATGGAGNACAGAGCGGCTTCCACTTCAAGATCAAGATNCTGCTAGTAGCGC 806
1408 TAGAGGTGCCCTTGAGAGACCGCGGTGTCACAGCCGGGACCGTCCCACTTCTACAGC 1467
807 TACGAAGTGCCTTGAGAGACCCACGCTGTGTGAGCAGGACCACTCCCACTTTACAGC 866
1468 GTGGAGTCTCCGAGAGCCCTTCGGGTGATGCTGCAACCGGAGTGCAGCGCGCGCTG 1527
867 GTGGAAATCTCAGAGGAAACCTTTGAGTGTGCTGTAGGAAGCTTGGTGGCGAGTG 926
1528 CTGCTGAACACAGCGGTGGCGCCCTGTCTTTGCGGACAGTTCCTTCAGCTGTCCACC 1587
927 TTGCTGAACACAGCGGTGGCGCCCTGTCTTCGCTGACACAGTTCCTGAGGTGTCCACT 986
1588 TCGTGCCTTCGAGPATATCAGGCTTCGCGAGCACCTCAGTCCCCTGATGTCTCAGC 1647
987 TCCCTGCCCTCCAGCACATCAGGCTTGGGGAACACCTCAGCCCACTATGCTCAGC 1046
1648 ACCAGCTGACAGGATCACCTGTGGAAACCGGACCTTGCGCCACGCCCGGTGGAAAC 1707
1047 ACCGATGGGCTCGTATACCTCTGGAAACCGGGAACACCACTCCGCAAGGTATCCAAAC 1106
1708 CTCTAGGGTCTCACCTTCTACCTGGCGCTGGAGAGCGGGGTGCGGACACAGGGGTG 1767
1107 CTCTAGGGTCTACATCTTCTACCTGGCACTGGAGGAGCGGTGGCTTGGCTCACGGTGT 1166
1768 TTCCTGCTAAACAGCAATGCCATGGATGTGCTGAGCGAGCGCTGCCCTTACGTGG 1827
1167 TTCCTGCTAAACAGCAATGCCATGGATGTATCTCTGCAACCCAGCGCCCACTAACCTGG 1226
1828 AGGTGACAGGTGGGATCTCGATGTCTACATCTTCTGGGCCACAGCGCGAGGCGTG 1887
1227 AGGTCAACGGGCGGGATCTCGATGTGTATGTCTTCTAGGCGCCACAGAGCCCAAGCGGT 1286

1888 GTGCAGCAGTACCTCGAAGTCTTGGGATACCGGTTTCATGCCGCCATCTATGGGGCCTGGGC 1947
1287 GTGCACAACTACTGATGTTTGGGATACCCCTTTCATGCTCTCATATGGGGCCTCGGC 1346
1948 TTCCACCTGTGCGCTGGGGCTTACTCTCCACACGCTATCACCCGACAGTGTGTGGAGAAC 2007
1347 TTCCACCTGTGCGCTGGGGCTTACTCTCCAGACCGCAATGTCCGCGAGGTAGTGGAGAAC 1406
2008 ATGACAGGGGCCACTTCCCTCGACGCTCCAAATGGACAGCAGCTGACATGACATCTCC 2067
1407 ATGACAGGACACACTTCCCTCGACGCTGCAATGGAATGACCTGACATGACAGCGC 1466
2068 CGGAGGAGCTTCAAGTTCGCTTCCGCGACTTCCCGGCCATGTGTGAGGAG 2127
1467 CGAAGAGACTTCACTTCAACAGGACAGACTTTCGCGACTTCCAGACATGTGTGCGGAG 1526
2128 CTGCAACAGGGCGCGCGCTTACATGATGATGCTGTGGATCTGTGCCATCAGCAGCTGGGC 2187
1527 CTGCAACAGGATGGCGGCTTACATGATGATGCTGTGGACCTGTCCATCAGCAGCGCAGC 1586
2188 CCTGCGGGAGCTACAGGCCCTACGACGAGGCTCTCGGAGGGGGTTCATCACCAAC 2247
1587 CTGCTGGAGTTACAGGCCCTTACGACGAGGCTCTGCGAGGGGCGTGTTCATCACCAAC 1646
2248 GAGACCGGCCAGCCCTGATTGGGAAGGTATGGCCCGGCTTCCACTTCCCTTCCCGACTTC 2307
1647 GAGACTGGGACGCGCTGATTGGGAAGTGTGGCCCTGGAAACCAACCGCTTCCCTGATTTC 1706
2308 ACCAACCCACAGCCCTGGCTGTGGGAGGACATGCTGTGGCTTCCACTTCCATCAGAGCTCTGAG 2427
1707 ACCAACCCCTGAGACCTTGAAGTGTGGGAGGACATGCTGTGCTGAGTTCACGCCAGGTG 1766
2368 CCCTTCGAGCGCTTGGATTGACATGACGAGCCTTCCAACTTCCATCAGAGCTCTGAG 2427
1767 CCCTTCGATGGCATGTGGCTCGACATGAACGAAACCGTCCAACTTCGTTAGAGGCTCTCAG 1826
2428 GACGCTGCCCCCAACAATGAGCTGGAGAACCCACCCCTACGCTCTGGGCTGTGGGGG 2487
1827 GAGGCTGCCCCCAACAATGAGCTGGAGAACCCCCCTATGTGCCGGGTGTGGCGG 1886
2488 ACCCTCAGGGGCGCACATCTGTGCTCAGACACAGTTCCTTCCACACACTACAAAC 2547
1887 ATCTTGCAGGAGCAGCAACCTGTGCTCCAGGCCAAATTCCTTCCACACACTACAAAC 1946
2548 CTGCAACACCTCTACGGCTTGACGAGCAGTCCGCTCCCAAGGCGCTGGTGAAGCT 2607
1947 CTCCAACACCTGTACGGCTTCACTGAAGCTATGCGCTCCAGCAGGGCCCTGGTCAAGCT 2006
2608 CGGGGACACGCCCAATTTGTGATCTCCGCTCGACCTTTGTGGCCACGGCCGATACGCC 2667
2007 CGGGGAACAGACCTTTGTGATCTCCGCTCAACCTTCTCGGGCCACGGCCGTTACGCT 2066
2668 GGCACCTGGAAGGGGACGTGTGAGCTCTGGGAGCAGCTGCTCTCCGTGCGCAGAA 2727
2067 GGTCACTGGAAGGGGATGTGCGGAGCTTGGGAGCATCTTGCATACTCTGTGCGCAGAC 2126
2728 ATCTGCAAGTTTAACTGCTGGGGTGTCTGCTGGGGCGGAGCTGTGGGCTTCTG 2787
2127 ATCTGCAAGTTTCAACCTCTGGGGTGTCCCTGTGGGGCGGACATCTGGGGCTTCATA 2186
2788 GGCACACCTCAGAGGAGCTGTGTGCTGCTGGACCCAGCTGGGGGCTTCTTACCCCTTC 2847
2187 GGAGACAGTCAAGAGAGCTGTGTGTGCTGGACCCAGTGGGGGCTTCTTACCCCTTC 2246
2848 ATGCGGAACCAACAGAGCTGCTGCTGCTGCCCGAGGAGCGGTACAGCTTACGAGCGG 2907
2247 ATGCGGAACCAATGACCTGAATAGCGTGTGCTCAGGAGCGGTACAGGTTACGAGGAG 2306
2908 GCCCAGAGGCGCATGAGGAAGCCCTCAGCTTGGCTGAGCAGCTCTCCCGCAGCTCTAC 2967
2307 GCGCAGAGGCGCATGAGGAGGCTTTCGCTTACGCTATGCTTCTGCGCTTCTGCTGTAC 2366


```
Db 2132 ATCTGAGTTCAACCTGCTGGCGTGCCTCGTGGGGGCGACATCTGCGGCTTCATA 2191
Qy 2788 GGCAACACCTCAGAGAGCTGTGTGCGCTGGACCCAGCTGGGGGCTTCTACCCCTTC 2847
Db 2192 GGAGACAGCTCAGAGAGCTGTGTGCGCTGGACCCAGCTGGGGGCTTCTACCCCTTC 2251
Qy 2848 ATCGGAACCAACACCTGCTGCTAGTCTGCCCGAGAGCCGTACAGCTTTCAGGAGCG 2907
Db 2252 ATCGGAACCAACACCTGCTGCTAGTCTGCCCGAGAGCCGTACAGCTTTCAGGAGCG 2311
Qy 2908 GCCAGAGCCATGAGGAGCCCTCACCTGGCTTACGACTCCCTCCCACTCTAC 2967
Db 2312 GCGAGAGCCATGAGGAGCCCTTCGCTTACGCTATGCCCTCTGTCCTTACCTGTAC 2371
Qy 2968 ACATGTTTCCACAGGCCACCTGCGGGGGAGACCGTGGCCCGGCGCCCTTCTTCTGGAG 3027
Db 2372 ACTCTTTCACCGCGCCACCTGTCAGAGAGACAGCGTGGCCCGGCGCCCTTCTTCTGGAG 2431
Qy 3028 TTCCCAAGAGCTCTAGCACTGTGAGCAACAGCTCTGTGGGGGGAGGCGCTG 3087
Db 2432 TTCCCTGAGGATCCAGCACTGTGTCGAGCGCCAGCTCTGTGGGGGGCGGCGCTG 2491
Qy 3088 CTATCACCCAGTCTTCCAGGCGGGAGGCGGAGCTGACTGCTACTTCCCTTGGGC 3147
Db 2492 CTATCACACCTGTGCTTGGAGCGCTGGGAAACTGAAGTGACGGGCTACTTCCCAAGGCG 2551
Qy 3148 ACATGTTTACACCTGACAGCGTGCCAAATAGAGCCCTTGGAGCCCTCCACCCCACT 3207
Db 2552 AGTGTGTAACAATGACAGTGTGTGAGTGTGCTGAGTCTCCCTGCTTCTTCTCATCA 2611
Qy 3208 GCAGCT---CCCCGTGAGCGAGCCATCCACAGGAGGGGAGTGGGTGACGCTGCCGGCC 3264
Db 2612 TCGGCTTCATCCTTTCAGATCTGCTGTCAGAGCAAGGGGAGTGGTGCACACTGGAAGCC 2671
Qy 3265 CCCTGGAACCAATCAAGTCTACCTCCGGGCTGGGTACATCATCCCTTCCGAGGGCCCT 3324
Db 2672 CCATCTGATACCAATCAAGTCTGAGGAGGAGGAGTACATCATCCCTGTCAGGGTCCC 2731
Qy 3325 GGCTTCAACCAACAGAGTCCCGCAGCAGCCATGCGCTGCTGGGCGCTTAACCAAG 3384
Db 2732 AGCTTCAACCAACAGAGTCCCGCAGCAGCCATGCGCTGCTGGGCGCTTAACCAAG 2791
Qy 3385 GTGGAGAGCCCGAGGGAGCTGTTCTGGGAGCATGGAGAGCCCTGGAAGTCTGGAG 3444
Db 2792 AGCGGAGGCCGATGGGAGCTGTTCTGGAGCAGCGGGAGAGCTCGGGTCTTGGAG 2851
Qy 3445 CGAGGGGCTTACACAGGTCATCTTCTGCGCAGGAGTAACAGGATCGTGAATGAGCTG 3504
Db 2852 CGTGGGGCTTACACAGTCTGCTTCTCAGCCCAAGCAATACCAATGTTGAACAAGTTA 2911
Qy 3505 GTACGTGTGACAGTGAAGGAGCTGGCTGCGAGTGCAGAGGTGACTGCTGGGGCTG 3564
Db 2912 GTGGGTGTGACCAAGAGGAGCTGAGCTAACTGAGGAGGTGACCGTCTTGGAGTG 2971
Qy 3565 GCCACGGCGCCAGCAGAGTCTCTTCCAAAGGTGCTCCCTGCTTCCCACTTCACTACAGC 3624
Db 2972 GCCACAGTCTTACCCAGGCTCTTCCAAAGGATCCCTGCTTCCCAATTTCACTACAGC 3031
Qy 3625 CCGACACCAAGTCTTGGACATCTGCTGCTGCTGTTGATGGAGGAGGATTTCTGCTC 3684
Db 3032 CTTGACAAAGAGCTGGCCATCCCTGCTCTACTGCTGATGGAGAGCTGTTTCAATC 3091
Qy 3685 AGCTGGT 3691
Db 3092 AGCTGGT 3098
```

RESULT 4

AY413854 2859 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes GAA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.

ACCESSION AY413854
VERSION AY413854.1 GI:39769816
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 2859)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrieri,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2859)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrieri,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1..2859
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
gene
<1..>2859
/gene="GAA"
/locus_tag="HCM5029"

Query Match 50.8%; Score 1881.6; DB 9; Length 2859;
Best Local Similarity 71.4%; Pred. No. 0;
Matches 1899; Conservative 0; Mismatches 760; Indels 0; Gaps 0;
Qy 1036 GAAACGGGTGCACACCCGGCGTCCAGAGAGTCCACACAGTGCAGCGTCCCGCC 1095
Db 199 GATGCCACGACACACCCCGCGCTCCAGAGAGTCCACACAGTGCAGCGTCCCGCC 258
Qy 1096 AACAGCGGCTTCGATTGCGCCCTGACAAAGGCCATCACCAGGAACATGCGAGCGCCG 1155
Db 259 AACAGCGGCTTCGATTGCGCCCGGACAGGCCATCACCAGGAACANNNNNNNNN 318
Qy 1156 GGCTGCTGTACATCCCTGCAAGCAGGCGGTGCGAGGAGCCAGATGGGCGAGCCCTGG 1215
Db 319 NNN 378
Qy 1216 TGCTTCTTCCACCCAGCTACCCAGCTGAGAACCTGAGCTCCTCTGAAATG 1275
Db 379 NNN 438
Qy 1276 GGCTACAGGCCACCCCTGACCCCGCTACACCCCGCCACCTTCTTCCCAAGGACATCCTGACC 1335
Db 439 NNN 498
Qy 1336 CTGCGGCTGAGCGTGATGATGGAGCTGAGAACCGGCTCCACTTCAGATCAAGATCCA 1395
Db 499 NNN 558
Qy 1396 GCTAACAGGCGCTACGAGTGCCTTGGAGACCCCGCGTGTCCACAGCGGCGCACCGTCC 1455
Db 559 GCTAACNNNGCTACGAGTGCCTTGGNNNNCCCGCGTGTCCACAGCGGCGCACCGTCC 618
Qy 1456 CCACTCTACAGCGTGGAGTTCCTCGAGGAGCCCTTCCGAGGAGCCCTTCCGCGGTGATCGTCACCGGCGAGCTG 1515
Db 619 CCACTCTNN 678
Qy 1516 GACGGCGCGTGTGCTGTGAACAGCAGCGTGGCGGCCCTGTTCTTCTTGGGACCAAGTTCCTT 1575


```
Db 184 CGGAACCAACAGCTGTCTCAGTCTGCCCGAGGCGGTACAGCTTCACGAGCGCGGCC 243
Qy 2911 CAGCAGGCATAGAGAGGCGCTCACCCTGCGGTAGCAGCTCTCCGCCACCTCTACACA 2970
Db 244 CAGCAGGCATAGAGAGGCGCTCACCCTGCGGTAGCAGCTCTCCGCCACCTCTACACA 303
Qy 2971 CTGTTCCACAGGCCCACTGCGGGGAGACCGTGGCCCGGCCCTCTTCCTGGAGTTC 3030
Db 304 CTGTTCCACAGGCCCACTGCGGGGAGACCGTGGCCCGGCCCTCTTCCTGGAGTTC 363
Qy 3031 CCCAAGACTCTAGCAGCTGAGCTGTGGACCAACAGCTCTGTGGGGAGGCGCTGCTC 3090
Db 364 CCCAAGACTCTAGCAGCTGAGCTGTGGACCAACAGCTCTGTGGGGAGGCGCTGCTC 423
Qy 3091 ATCACCACCACTGCTCCAGGCGGGAAGCGCAAGTGAAGTCTGCTTCCCTTGGGCACA 3150
Db 424 ATCACCACCACTGCTCCAGGCGGGAAGCGCAAGTGAAGTCTGCTTCCCTTGGGCACA 483
Qy 3151 TGGTAGCAGCTGTGAGAGCGGTGCAATAGAGGCGCTTGGAGAGCTCTCCACCCCACTGCA 3210
Db 484 TGGTAGCAGCTGTGAGAGCGGTGCAATAGAGGCGCTTGGAGAGCTCTCCACCCCACTGCA 543
Qy 3211 GCTCCCGGTGAGCAGCAGCATCCACAGGAGGCGAGTGGTGAAGTCTGCGGCGCCCTG 3270
Db 544 GCTCCCGGTGAGCAGCAGCATCCACAGGAGGCGAGTGGTGAAGTCTGCGGCGCCCTG 603
Qy 3271 GACACCATCAACGCTCCAGGCGGTGGGTACATCATCCCTCGAGGCGCTTGGCCCTC 3330
Db 604 GACACCATCAACGCTCCAGGCGGTGGGTACATCATCCCTCGAGGCGCTTGGCCCTC 663
Qy 3331 ACAACACAGAGTCCCGCCAGCAGCCATGGCCCTGCTGTGGCCCTTAACCAAGGTTGA 3390
Db 664 ACAACACAGAGTCCCGCCAGCAGCCATGGCCCTGCTGTGGCCCTTAACCAAGGTTGA 723
Qy 3391 GAGGCCGAGGAGCTGTCTGGAGCATGGAGAGAGCTGGAAGTCTGGAGCAGG 3450
Db 724 GAGGCCGAGGAGCTGTCTGGAGCATGGAGAGAGCTGGAAGTCTGGAGCAGG 783
Qy 3451 GCCTACACAGAGTCTCTTCCGCGCAGGAATAACACAGATCTGAATGAGTGTGACGT 3510
Db 784 GCCTACACAGAGTCTCTTCCGCGCAGGAAT-ACAGATCTGAATGAGTGTGACGT 842
Qy 3511 GTGACCACTGAGGAGCTGCGCTGACGTGACAGAGTGAATCTCTGCGCGTGGCCAG 3570
Db 843 GTGACCACTGAGGAGCTGCGCTGACGTGACAGAGTGAATCTCTGCGCGTGGCCAG 902
Qy 3571 GCGCCCGCAGAGTCTCTCCAAAGGTGCTCTGCTCCAACTTACCTACAGCCCGGAC 3630
Db 903 GCGCCCGCAGAGTCTCTCCAAAGGTGCTCTGCTCC-ACCTTACCTACANCCCGGAC 961
Qy 3631 ACCAAGTCTCTGACATCTGTGTCTCGCTGT--GATGGAGAGCAGTTTCTGTCAGCT 3688
Db 962 ACCAAGTCTCTGACATCTGTGTCTCGCTGTGTTGATGGAGAGCAATTTCTGTCACCT 1021
Qy 3689 GGTGTTGAC 3697
Db 1022 GGTGTTAAC 1030

RESULT 7
LOCUS CR617514
DEFINITION full-length cDNA clone CS0D1031YL16 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR617514
VERSION CR617514.1 GI:50498321
KEYWORDS HTC; CDSIT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1347)
```

```
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1347)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..1347
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1031YL16"
/tissue type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 25.9%; Score 957; DB 3; Length 1347;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Qy 63 CGGATCTTACTCGCGGAGAGAACCCAGCCCAAGCGTCTCCCGAAACGGAGTCCGAG 122
Db 97 CGGCAAGTACTCGCGGAGAGAACCCAGCCCAAGCGTCTCCCGAAACGGAGTCCGAG 156
Qy 123 GGAGTTCGATGGAGAGTTCGAACAGCTGTGGGAGAGGCCCGAGCTGATCTTCC 182
Db 157 GGAGTTCGATGGAGAGTTCGAACAGCTGTGGGAGAGGCCCGAGCTGATCTTCC 216
Qy 183 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAATCGC 242
Db 217 TCCGCTGAGGCTGGCGGAGCTCCAGCTGATCTGAAGATACAGGAGAGGAGCGAATCGC 276
Qy 243 CTGGAAGAACTAAAGCTTTGACGGCTTGACGAAAGTGGGAGAGGAGAGCGAGTCTAT 302
Db 277 CTGGAAGAACTAAAGCTTTGACGGCTTGACGAAAGTGGGAGAGGAGAGCGAGTCTAT 336
Qy 303 AGCAACCTCAATGTCATCTTGGCCAGTATGGTCTGGAGCGGAGAGGAGCGCTCGGCA 362
Db 337 AGCAACCTCAATGTCATCTTGGCCAGTATGGTCTGGAGCGGAGAGGAGCGCTCGGCA 396
Qy 363 GGTGACCAAGCTCCCTCAGTGGACCCAGGAGAGCGGCTGGATGACCCAGGCTGGA 422
Db 397 GGTGACCAAGCTCCCTCAGTGGACCCAGGAGAGCGGCTGGATGACCCAGGCTGGA 456
Qy 423 AAAGCTGTGGCAAGGCGGAGACCTCTCGGAAATTTCTCGGCGAAGAACTGGACAAGCT 482
Db 457 AAAGCTGTGGCAAGGCGGAGACCTCTCGGAAATTTCTCGGCGAAGAACTGGACAAGCT 516
Qy 483 CTGGCGGAGTTCCTGTCATCACAAAGAGAAAGTTTACGAGTACAACTCTGCTGGAGAC 542
Db 517 CTGGCGGAGTTCCTGTCATCACAAAGAGAAAGTTTACGAGTACAACTCTGCTGGAGAC 576
Qy 543 CTGTAGCAGGACCGGAAGAAATCCAGAGAACTGCTTAGCCCTCGGAGCTGAGCGACAT 602
Db 577 CTGTAGCAGGACCGGAAGAAATCCAGAGAACTGCTTAGCCCTCGGAGCTGAGCGACAT 636
Qy 603 CAAGGCGAGCGTCTCTGCAAGCAGGAGCAGCGAGCTGGAAGAGAGTCTGCGAGCATCAA 662
Db 637 CAAGGCGAGCGTCTCTGCAAGCAGGAGCAGCGAGCTGGAAGAGAGTCTGCGAGCATCAA 696
Qy 663 CCAGGCGCTGGACCGGCTGCGAGGCTGAGCCACGAGGCTACAGCACTGAGGCTGAGT 722
```

Db	697	CCAGGGCTTGGACCGCCTGCGCAGGGT	CAGCCACGAGGGCTTACAGCACTGAGGCTGAGTT	756
QY	723	CGAGGAGCCCGAGGCTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACTCACGGACAA		782
Db	757	CGAGGAGCCCGAGGCTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACTCACGGACAA		816
QY	783	GGAGCTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCACAA		842
Db	817	GGAGCTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAAGCACAA		876
QY	843	CCACTACCAAGACAGCTGGAGATTGGCCACGAGAGAGCTGAGGCACGAGAGAGCGTGGG		902
Db	877	CCACTACCAAGACAGCTGGAGATTGGCCACGAGAGAGCTGAGGCACGAGAGAGCGTGGG		936
QY	903	CGAGCGGAGCGTGTGAGCGCGCAGCGAGAGCAGCGCCCTGCTGGAGGGCGGACCAA		962
Db	937	CGAGCGGAGCGTGTGAGCGCGCAGCGAGAGCAGCGCCCTGCTGGAGGGCGGACCAA		996
QY	963	GGAGCTGGGCTACACGGTGAAGAAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC		1022
Db	997	GGAGCTGGGCTACACGGTGAAGAAGCATCTGCAGGACCTGTCCGGCAGGATCTCCAGAGC		1056
QY	1023	TCGCGCCGAGGCA	1035	
Db	1057	TCGGCACAAACGAA	1069	
RESULT 8				
LOCUS	CR603866	1373 bp mRNA linear	HTC 21-JUL-2004	
DEFINITION	full-length cDNA clone CS0D1004YE06 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR603866			
VERSION	CR603866.1	GI:50484673		
KEYWORDS	HTC; CNSLT_cDNA.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished			
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue			
REFERENCE	2 (bases 1 to 1373)			
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.			
FEATURES	Location/Qualifiers			
source	1..1373			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="CS0D1004YE06"			
	/tissue_type="Placenta Cot 25-normalized"			
	/plasmid="pCMVSPORT_6"			
ORIGIN				
Query Match	25.9%; Score 957; DB 3; Length 1373;			
Best Local Similarity	99.0%; Pred. No. 1.2e-185;			
Matches 963; Conservative	0; Mismatches 10; Indels 0; Gaps 0;			
QY	63	CGGATCCTACTCGCGGAGAGAACCGCCAGCGCTCCCGAAGCGGAGTCCGGAGA	122	
RESULT 9				
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388	1385 bp mRNA linear	HTC 21-JUL-2004	
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1011YN09 of Placenta Cot 25-normalized of Homo sapiens (human).			
ACCESSION	CR619388			
LOCUS	CR619388			
DEFINITION	full-length cDNA clone CS0D1			


```
VERSION CR619388.1 GI:50500195
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1385)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1385)
Direct Submission
Genoscope.
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..1385
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1011YN09"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 25.98; Score 957; DB 3; Length 1385;
Best Local Similarity 99.08; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 63 CGGATCTCTACTCGGGGAGAGAACACAGCCCAAGCCGTCCTCCCGAAACGCGAGTCCGGAGA 122
DB 90 CGGCAGTACTCGGGGAGAGAACACAGCCCAAGCCGTCCTCCCGAAACGCGAGTCCGGAGA 149
QY 123 GAGATTCCCGCATGGAGAGTTGAACACAGCTGTGGGAGAGGCCCGACGACTGCACTTCC 182
DB 150 GGAGTTCCCGCATGGAGAGTTGAACACAGCTGTGGGAGAGGCCCGACGACTGCACTTCC 209
QY 183 TCCCGTAGGCTGGCGAGCTCCACCTGATCTGAGATACAGAGAGGAGGAGCAACTCCG 242
DB 210 TCCCGTAGGCTGGCGAGCTCCACCTGATCTGAGATACAGAGAGGAGGAGCAACTCCG 269
QY 243 CTGGAAGAACTAAAGCTTGACGGCTTGACGAGATGGGAGAGGAGGAGGAGCAACTCAT 302
DB 270 CTGGAAGAACTAAAGCTTGACGGCTTGACGAGATGGGAGAGGAGGAGGAGCAACTCAT 329
QY 303 ACCGAACCTCAATGTATCTTGGCCCAAGTATGTCTGGACGGAAGAGAGCACTCGGCA 362
DB 330 ACCGAACCTCAATGTATCTTGGCCCAAGTATGTCTGGACGGAAGAGAGCACTCGGCA 389
QY 363 GGTGACGAGCACTCCCTAGTGGCCACCCAGGAGAGCGGCTGGATGACCCCGGCTGGA 422
DB 390 GGTGACGAGCACTCCCTAGTGGCCACCCAGGAGAGCGGCTGGATGACCCCGGCTGGA 449
QY 423 AAGCTGTGGCACAAGCGAAGACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAAGCT 482
DB 450 AAGCTGTGGCACAAGCGGAGAACCTCTGGGAAATTTCTCGGCGAAGAACTGGACAAAGCT 509
QY 483 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTTCAAGATCAACAGCTCTCTGCTGGAGAC 542
DB 510 CTGGCGGGAGTTCTCTGCATCAAAAGAGAAAGTTTCAAGATCAACAGCTCTCTGCTGGAGAC 569
QY 543 CTTGAGCAGGACCGAAGAAATTCACGAGAAAGCTCATTTAGCCCTCGGACCTGAGGACAT 602
DB 570 CTTGAGCAGGACCGAAGAAATTCACGAGAAAGCTCATTTAGCCCTCGGACCTGAGGACAT 629
```

```
QY 603 CAAAGGCGAGCGTCTCTGCACAGCAGGCGACACGCGAGCTGAAGGAGAGCTGCGCAGCATCAA 662
DB 630 CAAAGGCGAGCGTCTCTGCACAGCAGGCGACACGCGAGCTGAAGGAGAGCTGCGCAGCATCAA 689
QY 663 CAAAGGCGAGCGTCTCTGCACAGGCGTCAAGGAGCTGAGGAGAGCTGAGGAGCTGAGTT 722
DB 690 CAAAGGCGAGCGTCTCTGCACAGGCGTCAAGGAGCTGAGGAGAGCTGAGGAGCTGAGTT 749
QY 723 CAAAGGAGCCAGGCGTGAATGACCTGTGGGACCTTGGGCGAGTCCGCAACCTCAACGACAA 782
DB 750 CAAAGGAGCCAGGCGTGAATGACCTGTGGGACCTTGGGCGAGTCCGCAACCTCAACGACAA 809
QY 783 GGAGCTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAGACAA 842
DB 810 GGAGCTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTTCGAAGCCAAATCGAGAGACAA 869
QY 843 CCNCTACCAAGAGCAGCTGGAGATTTCGCACGAGAGAGCTGAGGAGAGAGAGCGTGGG 902
DB 870 CCNCTACCAAGAGCAGCTGGAGATTTCGCACGAGAGAGCTGAGGAGAGAGAGCGTGGG 929
QY 903 CGACGCGGAGCGTGTGAGCGCGAGCGCGAGAGAGACGCGCTCTCGAGGCGGCGACAA 962
DB 930 CGACGCGGAGCGTGTGAGCGCGAGCGCGAGAGAGAGCGCGCTCTCGAGGCGGCGACAA 989
QY 963 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTTGTCCGGCAGGATCTCCAGAGC 1022
DB 990 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGAGACCTGTCCGGCAGGATCTCCAGAGC 1049
QY 1023 TCGCGCCGAGGCA 1035
DB 1050 TCGGCACAACGAA 1062
```

```
RESULT 10
CR606004
LOCUS DEFINITION
CR606004 full-length cDNA clone CS0D1030YG03 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR606004.1 GI:50486811
VERSION CR606004.1
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1392)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1392)
Direct Submission
Genoscope.
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..1392
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1030YG03"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
```

```
Query Match      25.9%; Score 957; DB 3; Length 1392;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTTACTCGCGGAGAGAACACAGCCCAAGCCGTCGCCGAACCGGAGTCCGGAGA 122
DB 41 CGGAGAGTACTCGCGGAGAGAACACAGCCCAAGCCGTCGCCGAACCGGAGTCCGGAGA 100
QY 123 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCACGAGCTGATCTTCC 182
DB 101 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCACGAGCTGATCTTCC 160
QY 183 TCCCGTAGGCTGCGCGAGCTCCACGCTGTATCTGAAGATACAGGAGAGGAGCAACTCGC 242
DB 161 TCCCGTAGGCTGCGCGAGCTCCACGCTGTATCTGAAGATACAGGAGAGGAGCAACTCGC 220
QY 243 CTGGAAGAACTAAAGCTTGACCGCTTGACGAGATGGGAGAGAGGAGAGGAGAGTCAAT 302
DB 221 CTGGAAGAACTAAAGCTTGACCGCTTGACGAGATGGGAGAGAGGAGAGGAGAGTCAAT 280
QY 303 ACGCAACCTCAATGTCTATCTTGCCCAAGTATGTCTGGACGGAAGAGAGAGCTCGGCA 362
DB 281 ACGCAACCTCAATGTCTATCTTGCCCAAGTATGTCTGGACGGAAGAGAGAGCTCGGCA 340
QY 363 GGTGACGAGCAACTCCCTCAGTGGCAACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGA 422
DB 341 GGTGACGAGCAACTCCCTCAGTGGCAACCCAGGAAGACGGGCTGGATGACCCCAAGGCTGA 400
QY 423 AAGCTGTGGCACAAGCGAGACCTCTGGGAATTTCTCGGCGAAGAACTGGACAAGCT 482
DB 401 AAGCTGTGGCACAAGCGAGACCTCTGGGAATTTCTCGGCGAAGAACTGGACAAGCT 460
QY 483 CTGCGCGGAGTCTCTGATCACAAGAGAGAAAGTTTACAGATCAACAGTCTCTGTGGAGAC 542
DB 461 CTGCGCGGAGTCTCTGATCACAAGAGAGAAAGTTTACAGATCAACAGTCTCTGTGGAGAC 520
QY 543 CTTGAGCAGGACCGAAGAAATCCACAGAGAACCGTATTAGCCCTCGGACCTTGAGCGACAT 602
DB 521 CTTGAGCAGGACCGAAGAAATCCACAGAGAACCGTATTAGCCCTCGGACCTTGAGCGACAT 580
QY 603 CAGGCGAGGCTCTGACAGGAGGACACAGGAGCTGAAGAGAGAGTCTGGGAGAGTCAAA 662
DB 581 CAGGCGAGGCTCTGACAGGAGGACACAGGAGCTGAAGAGAGAGTCTGGGAGAGTCAAA 640
QY 663 CAGGCGCTGAGCGGCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 722
DB 641 CAGGCGCTGAGCGGCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 700
QY 723 CGAGGAGCCAGGCTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACCTTCAAGCAAA 782
DB 701 CGAGGAGCCAGGCTGATTGACCTGTGGGACCTGGCGCAGTCCGCCAACCTTCAAGCAAA 760
QY 783 GGAGCTGAGGCGCTTCGGGAGAGCTCAAGCACTTGAAGCCAAATCGAGAGACAA 842
DB 761 GGAGCTGAGGCGCTTCGGGAGAGCTCAAGCACTTGAAGCCAAATCGAGAGACAA 820
QY 843 CCACTACCAAGACAGCTGAGAGTTGCGCACAGAGAGCTGAGGACGACAGAGCGTGGG 902
DB 821 CCACTACCAAGACAGCTGAGAGTTGCGCACAGAGAGCTGAGGACGACAGAGCGTGGG 880
QY 903 CGACGCGAGCGTGTGAGCGGACCGCGAGAGACAGCCCTGTCTGGAGGGGCGGACAA 962
DB 881 CGACGCGAGCGTGTGAGCGGACCGCGAGAGACAGCCCTGTCTGGAGGGGCGGACAA 940
QY 963 GGAGCTGGGCTACACGCTGAAGAGCATCTGAGGAGCCTGTCCGCGAGGATCTCCAGAGC 1022
DB 941 GGAGCTGGGCTACACGCTGAAGAGCATCTGAGGAGCCTGTCCGCGAGGATCTCCAGAGC 1000
QY 1023 TCGCGCGGAGGCA 1035
DB 1001 TCGGCAACAGAA 1013
```

```
RESULT 11
CR596948
LOCUS
DEFINITION
full-length cDNA clone CS0DI035YA24 of Placentia Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR596948
VERSION
CR596948.1 GI:50477755
KEYWORDS
HTC; CNSLT CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1399)
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1399)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1399
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI035YA24"
/tissue_type="Placentia Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
```

```
Query Match      25.9%; Score 957; DB 3; Length 1399;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTTACTCGCGGAGAGAACACAGCCCAAGCCGTCGCCGAACCGGAGTCCGGAGA 122
DB 88 CGGCAAGTACTCGCGGAGAGAACACAGCCCAAGCCGTCGCCGAACCGGAGTCCGGAGA 147
QY 123 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCACGAGCTGATCTTCC 182
DB 148 GGAGTTCCGATCGAGAGTTGAACACAGCTGTGGGAGAGAGCCACGAGCTGATCTTCC 207
QY 183 TCCCGTAGGCTGCGCGAGCTCCACGCTGTGAAGATACAGGAGAGGAGCAACTCGC 242
DB 208 TCCCGTAGGCTGCGCGAGCTCCACGCTGTGAAGATACAGGAGAGGAGCAACTCGC 267
QY 243 CTGGAAGAACTAAAGCTTGACCGCTTGGAAGAGAGGAGAGGAGGAGGAGTCAAT 302
DB 268 CTGGAAGAACTAAAGCTTGACCGCTTGGAAGAGAGGAGAGGAGGAGTCAAT 327
QY 303 ACGCAACCTCAATGTCTATCTTGCCCAAGTATGTCTGGAGAGAGAGAGAGAGCTCGGCA 362
DB 328 ACGCAACCTCAATGTCTATCTTGCCCAAGTATGTCTGGAGAGAGAGAGAGAGCTCGGCA 387
QY 363 GGTGACGAGCAACTCCCTCAGTGGCAGCCACGAGAGAGGAGGAGTGGATGAGCCAGGCTGA 422
DB 388 GGTGACGAGCAACTCCCTCAGTGGCAGCCACGAGAGAGGAGGAGTGGATGAGCCAGGCTGA 447
QY 423 AAGCTGTGGCACAAGCGAGAGACCTCTGGGAAATTTCTCCGCGAGAGAACTCGACAAGCT 482
DB 448 AAGCTGTGGCACAAGCGAGAGACCTCTGGGAAATTTCTCCGCGAGAGAACTCGACAAGCT 507
```

```
QY 483 CTGGCGGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAGAGTACAAAGTCTCTGCTGGAGAC 542
Db 508 CTGGCGGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAGAGTACAAAGTCTCTGCTGGAGAC 567
QY 543 CTGAGCAGGACCGAGAGAAATCCACAGAGAACGTCATTAGCCCTCGGACCTGAGCGACAT 602
Db 568 CTTGAGCAGGACCGAGAGAAATCCACAGAGAACGTCATTAGCCCTCGGACCTGAGCGACAT 627
QY 603 CAAAGGCGAGCGTCTCTGCACAGCAGGACACAGCGAGCTCAAGGAGAAAGTCTGCGAGCATCAA 662
Db 628 CAAAGGCGAGCGTCTCTGCACAGCAGGACACAGCGAGCTCAAGGAGAAAGTCTGCGAGCATCAA 687
QY 663 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 722
Db 688 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 747
QY 723 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 782
Db 748 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 807
QY 783 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 842
Db 808 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 867
QY 843 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 902
Db 868 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 927
QY 903 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 962
Db 928 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 987
QY 963 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 1022
Db 988 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 1047
QY 1023 TCAGCGCGAGGCA 1035
Db 1048 TCAGCGCGAGGCA 1060

RESULT 12
LOCUS CR625424
DEFINITION full-length cDNA clone CS0DJ007YP06 of T cells (Jurkat cell line)
ACCESSION CR625424
VERSION 1
KEYWORDS full-length cDNA libraries and normalization
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1399)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1399)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
Location/Qualifiers
1. .1399
source
```

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ007YP06"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"
```

ORIGIN

```
Query Match 25.9%; Score 957; DB 3; Length 1399;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCTTACTCGCGGAGAGAAACAGCCCAAGCGCTCCCGAAACGCGAGTCCGGAGA 122
Db 92 CGGCAAGTACTCGCGGAGAGAAACAGCCCAAGCGCTCCCGAAACGCGAGTCCGGAGA 151
QY 123 GGAGTTCCGATGAGAGAGTTGAACACAGCTGTGGAGAGAGGCCAGCGACTGCATCTTCC 182
Db 152 GGAGTTCCGATGAGAGAGTTGAACACAGCTGTGGAGAGAGGCCAGCGACTGCATCTTCC 211
QY 183 TCCCGTGGAGGCTGGCGGAGCTCCAGCGTGTGAACAGATACAGGAGAGGAGCGAACTCGC 242
Db 212 TCCCGTGGAGGCTGGCGGAGCTCCAGCGTGTGAACAGATACAGGAGAGGAGCGAACTCGC 271
QY 243 CTGGAAGAAACTAAAGCTTTGACGGTTTGAACGAAATGGGGAGAGAAAGCGAGACTCAT 302
Db 272 CTGGAAGAAACTAAAGCTTTGACGGTTTGAACGAAATGGGGAGAGAAAGCGAGACTCAT 331
QY 303 AGCAACCTCAATGTCTTGGCCAGTATGGTCTGGACGAAAGAAAGAGAGCGCTCGGCA 362
Db 332 AGCAACCTCAATGTCTTGGCCAGTATGGTCTGGACGAAAGAAAGAGAGCGCTCGGCA 391
QY 363 GTGTACCAGCAACTCCCTCAGTGGCACCCAGGAAAGACGGGCTGGATGACCCAGAGTGA 422
Db 392 GTGTACCAGCAACTCCCTCAGTGGCACCCAGGAAAGACGGGCTGGATGACCCAGAGTGA 451
QY 423 AAAGCTGTGGCAACAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTGGACAAGCT 482
Db 452 AAAGCTGTGGCAACAAGGCGAAGACCTCTGGGAAATTTCTCCGCGAAGAACTGGACAAGCT 511
QY 483 CTGGCGGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAGAGTACAAAGTCTCTGCTGGAGAC 542
Db 512 CTGGCGGAGTTCTCTGCATCACAAGAGAGAAAGTTTCAGAGTACAAAGTCTCTGCTGGAGAC 571
QY 543 CTTGAGCAGGACCGAGAGAAATCCACAGAGAACGTCATTAGCCCTCGGACCTGAGCGACAT 602
Db 572 CTTGAGCAGGACCGAGAGAAATCCACAGAGAACGTCATTAGCCCTCGGACCTGAGCGACAT 631
QY 603 CAAAGGCGAGCGTCTCTGCACAGCAGGACACAGCGAGCTCAAGGAGAAAGTCTGCGAGCATCAA 662
Db 632 CAAAGGCGAGCGTCTCTGCACAGCAGGACACAGCGAGCTCAAGGAGAAAGTCTGCGAGCATCAA 691
QY 663 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 722
Db 692 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 751
QY 723 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 782
Db 752 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 811
QY 783 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 842
Db 812 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 871
QY 843 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 902
Db 872 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 931
QY 903 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 962
Db 932 CAAAGGCGAGCGTCTCTGCACAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGT 991
```

```
QY 963 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTTCGGCAGGATCTCCAGAGC 1022
DB 992 GGAGCTGGGCTACACGGTGAAGAGCATCTGCAGGACCTGTTCGGCAGGATCTCCAGAGC 1051
QY 1023 TCGCGCCGAGGCA 1035
DB 1052 TCGGCACAACGAA 1064

RESULT 13
CR591537
LOCUS
DEFINITION
Full-length cDNA clone CS0D005YL06 of T cells (Jurkat cell line)
Cot 10-normalized of Homo sapiens (human).
CR591537
ACCESSION
VERSION
CR591537.1 GI:50472344
KEYWORDS
HTC; CNSLT_cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1402)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
REMARK
2 (bases 1 to 1402)
Genoscope.
Direct Submission
AUTHORS
TITLE
JOURNAL
REMARK

REFERENCE
1. 1402
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1. 1402
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D005YL06"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 25.9%; Score 957; DB 3; Length 1402;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCCTACTCGCGGAGAGAACACAGCCCAAGCGTCCCGAAACGGAGTCCGGAGA 122
DB 76 CGGCAAGTACTCGCGGAGAGAACACAGCCCAAGCGTCCCGAAACGGAGTCCGGAGA 135
QY 123 GGAGTTCCCGATGGAGAACTTGACACAGCTGTGGGAGAGGCCCGGACTGCATCTTCC 182
DB 136 GGAGTTCCCGATGGAGAACTTGACACAGCTGTGGGAGAGGCCCGGACTGCATCTTCC 195
QY 183 TCCGTTGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGGAGGAGCGAACTCGC 242
DB 196 TCCGTTGAGGCTGGCCGAGCTCCACGCTGATCTGAAGATACAGGAGGAGCGAACTCGC 255
QY 243 CTGGAAGAAACTAAAGCTTGAACGGCTTGGACGAGATGGGGAGAGAAAGCGAGACTCAT 302
DB 256 CTGGAAGAAACTAAAGCTTGAACGGCTTGGACGAGATGGGGAGAGAAAGCGAGACTCAT 315
QY 303 ACGCAACCTCAATGTATCTTGGCCAAAGTATGCTTGGACGGAAGAAAGCGCTCGGCA 362
DB 316 ACGCAACCTCAATGTATCTTGGCCAAAGTATGCTTGGACGGAAGAAAGCGCTCGGCA 375
```

```
QY 363 GGTGACCAACGAACTCCCTCAGTGGCACCAGGAAGACGCGGTGGATGACCCCGAGGCTGA 422
DB 376 GGTGACCAACGAACTCCCTCAGTGGCACCAGGAAGACGCGGTGGATGACCCCGAGGCTGA 435
QY 423 AAAGCTGTGGCAAGGGCGAGACCTCTGGGAAATTTCTCCGGCGGAAGAACTGGACNAGCT 482
DB 436 AAAGCTGTGGCAAGGGCGAGACCTCTGGGAAATTTCTCCGGCGGAAGAACTGGACNAGCT 495
QY 483 CTGGCGGAGTTCTTCGTCATCAAAAGAGAAAGTTCACAGTACAAAGCTCTCTCTGGAGAC 542
DB 496 CTGGCGGAGTTCTTCGTCATCAAAAGAGAAAGTTCACAGTACAAAGCTCTCTCTGGAGAC 555
QY 543 CTTGAGCAGGACCGAAGAAATTCACAGAAACGTCATTAGCCCTCGGACCTCGAGCGACAT 602
DB 556 CTTGAGCAGGACCGAAGAAATTCACAGAAACGTCATTAGCCCTCGGACCTCGAGCGACAT 615
QY 603 CAAAGGCGAGCTCTTCGTCACAGCAGGACACAGGAGCTGAAGGAGAGCTGCGCAGATCAA 662
DB 616 CAAAGGCGAGCTCTTCGTCACAGCAGGACACAGGAGCTGAAGGAGAGCTGCGCAGATCAA 675
QY 663 CCAGGCGCTGGACCGCCCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 722
DB 676 CCAGGCGCTGGACCGCCCTGCGCAGGCTCAGCCACAGGCTACAGCACTGAGGCTGAGTT 735
QY 723 CGAGGAGCCCGAGGCTGATTGACCTCTGGGACCTGCGCAGCTCCGCAACCTCACGGACAA 782
DB 736 CGAGGAGCCCGAGGCTGATTGACCTCTGGGACCTGCGCAGCTCCGCAACCTCACGGACAA 795
QY 783 GGAGCTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTCGAAGCAAAATTCGAAGACGACAA 842
DB 796 GGAGCTGGAGGCGTTCCGGGAGGAGCTCAAGCACTTCGAAGCAAAATTCGAAGACGACAA 855
QY 843 CCACCTACAGAGCAGCTGGAGATTGCCACAGAGAGCTGAGGCACGACGAGAGCGTGGG 902
DB 856 CCACCTACAGAGCAGCTGGAGATTGCCACAGAGAGCTGAGGCACGACGAGAGCGTGGG 915
QY 903 CGACGCGAGCGTGTGAGCCGCGAGAGCAACGCCCTCTGTGGAGGGGCGGACCA 962
DB 916 CGACGCGAGCGTGTGAGCCGCGAGAGCAACGCCCTCTGTGGAGGGGCGGACCA 975
QY 963 GGAGCTGGGCTACAGGCTGAGAGAGCATCTGCAGGACCTGTCGGCAGGATCTCCAGAGC 1022
DB 976 GGAGCTGGGCTACAGGCTGAGAGAGCATCTGCAGGACCTGTCGGCAGGATCTCCAGAGC 1035
QY 1023 TCGCGCCGAGGCA 1035
DB 1036 TCGGCACAACGAA 1048
```

```
RESULT 14
CR619887
LOCUS
DEFINITION
full-length cDNA clone CS0DE003YG19 of Placenta of Homo sapiens
(human).
CR619887
ACCESSION
VERSION
CR619887.1 GI:50500694
KEYWORDS
HTC; CNSLT_cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1409)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
REMARK
2 (bases 1 to 1409)
Genoscope.
Direct Submission
AUTHORS
TITLE
JOURNAL
REMARK

REFERENCE
1 (bases 1 to 1409)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
```

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source Location/Qualifiers
1..1409
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DE003YB19"
/tissue_type="Placenta"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 25.9%; Score 957; DB 3; Length 1409;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCCTACTCCGGGAGAGAACAGCCAGCCCGTCCCGAAGCGGAGTCCGGAGA 122
DB |||||
QY 92 CGCAGAGTACTCGGGAGAGAGAACAGCCCAAGCCGTCCTCCGAAAGCGAGTCCGGAGA 151
DB |||||
QY 123 GGAGTTCGGCATCGAAGGTGAACACAGCTGTGGGAGAGGCCCGACGACTGCACTTCC 182
DB |||||
QY 152 GGAGTTCGGCATCGAAGGTGAACACAGCTGTGGGAGAGGCCCGACGACTGCACTTCC 211
DB |||||
QY 183 TCCCGTAGGCTGGCGAGCTCCAGCTGATCTGAAGATACAGGAGGAGGAGCAATCCG 242
DB |||||
QY 212 TCCCGTAGGCTGGCGAGCTCCAGCTGATCTGAAGATACAGGAGGAGGAGCAATCCG 271
DB |||||
QY 243 CTGGAAGAACTAAAGCTTGACGGCTTGACGAAGATGGGAGAGGAGGAGCAATCCAT 302
DB |||||
QY 272 CTGGAAGAACTAAAGCTTGACGGCTTGACGAAGATGGGAGAGGAGGAGCAATCCAT 331
DB |||||
QY 303 AGCAACCTCAATGTCTATCTTGCCCAAGTATGCTTGGAAGGAGGAGGAGCAATCCGCA 362
DB |||||
QY 332 AGCAACCTCAATGTCTATCTTGCCCAAGTATGCTTGGAAGGAGGAGGAGCAATCCGCA 391
DB |||||
QY 363 GGTGACAGCACTCCTCAGTGGCCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 422
DB |||||
QY 392 GGTGACAGCACTCCTCAGTGGCCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 451
DB |||||
QY 423 AAAGCTGTGCAAGCGAAGACCTCTGGGAAATCTCGGGAAGAACTGGGAGCAAGCT 482
DB |||||
QY 452 AAAGCTGTGCAAGCGGAGACCTCTGGGAAATCTCGGGAAGAACTGGGAGCAAGCT 511
DB |||||
QY 483 CTGGCGGAGTTCCTGCATCAAAAGAGAAAGTTCAAGATCAACGTCCTGCTGGAGAC 542
DB |||||
QY 512 CTGGCGGAGTTCCTGCATCAAAAGAGAAAGTTCAAGATCAACGTCCTGCTGGAGAC 571
DB |||||
QY 543 CTTGACAGGAGCCGAGAAATCAACAGAACGTCATTAGCCCTCCGAGCTGAGGACAT 602
DB |||||
QY 572 CTTGACAGGAGCCGAGAAATCAACAGAACGTCATTAGCCCTCCGAGCTGAGGACAT 631
DB |||||
QY 603 CAAGGCGAGCTCCTGCACAGGAGGAGCAGGAGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 662
DB |||||
QY 632 CAAGGCGAGCTCCTGCACAGGAGGAGCAGGAGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAG 691
DB |||||
QY 663 CCAAGGCGCTGGACCGCTCGGAGGAGTCAAGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 722
DB |||||
QY 692 CCAAGGCGCTGGACCGCTCGGAGGAGTCAAGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 751
DB |||||
QY 723 CGAGGAGCCGAGGATGATGACCTGTGGAGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 782
DB |||||
QY 752 CGAGGAGCCGAGGATGATGACCTGTGGAGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 811
DB |||||
QY 783 GGAGCTGAGGAGGCTCCGAGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAGGAGGAGGAG 842
DB |||||
QY 812 GGAGCTGAGGAGGCTCCGAGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAGAGGAGGAGGAG 871
DB |||||

QY 843 CCCTACACAGAGCAGCTGGAGATTCCGACAGAGAGCTTGAGCAGCAGCAGAGAGCGTGGG 902
DB |||||
QY 872 CCCTACACAGAGCAGCTGGAGATTCCGACAGAGAGCTTGAGCAGCAGCAGAGAGCGTGGG 931
DB |||||
QY 903 CGACGCGGAGCTGTGAGCGGAGCGGAGAGGAGCAGCCCTGCTGCGAGGCGGAGCAGCAG 962
DB |||||
QY 932 CGACGCGGAGCTGTGAGCGGAGCGGAGAGGAGCAGCCCTGCTGCGAGGCGGAGCAGCAG 991
DB |||||
QY 963 GGAGCTGGGCTACACGGTGAAGAAGCATCTGCGAGGACCTGTGCGGAGGAGTCTCCAGAGC 1022
DB |||||
QY 992 GGAGCTGGGCTACACGGTGAAGAAGCATCTGCGAGGACCTGTGCGGAGGAGTCTCCAGAGC 1051
DB |||||
QY 1023 TCGCGCCGAGGCA 1035
DB |||||
QY 1052 TCGGCACAAACGAA 1064
DB |||||

RESULT 15
CR607197
LOCUS 1411 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DI023YB19 of Placenta Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR607197
VERSION CR607197.1 GI:50488004
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1411)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normaliza-
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 1411)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source Location/Qualifiers
1..1411
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI023YB19"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 25.9%; Score 957; DB 3; Length 1411;
Best Local Similarity 99.0%; Pred. No. 1.2e-185;
Matches 963; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 63 CGGATCCTACTCCGGGAGAGAACAGCCAGCCCGTCCCGAAGCGGAGTCCGGAGA 122
DB |||||
QY 90 CGGCAAGTACTCGGGGAGAGAACAGCCCAAGCCGTCCTCCGAAAGCGAGTCCGGAGA 149
DB |||||
QY 123 GGAGTTCGGCATCGAAGGTGAACACAGCTGTGGGAGAGGCCCGACGACTGCACTTCC 182
DB |||||
QY 150 GGAGTTCGGCATCGAAGGTGAACACAGCTGTGGGAGAGGCCCGACGACTGCACTTCC 209
DB |||||
QY 183 TCCCGTAGGCTGGCGAGCTCCAGCTGATCTGAAGATACAGGAGGAGGAGGAGGAGGAGGAG 242
DB |||||
QY 210 TCCCGTAGGCTGGCGAGCTCCAGCTGATCTGAAGATACAGGAGGAGGAGGAGGAGGAGGAG 269
DB |||||

```
QY 243 CTGGAAGAACTAAAGCTTGA CGGCTTGA CGAAGATGGGAGAAAGAAAGCGAGACTCAT 302
Db |||||||
QY 270 CTGGAAGAACTAAAGCTTGA CGGCTTGA CGAAGATGGGAGAAAGAAAGCGAGACTCAT 329
Db |||||||
QY 303 AGCAACCTCAATGTCTATCTTGGCCAAAGTATGGTCTGGACGGNAAGAAAGACGCTCGGCA 362
Db |||||||
QY 330 AGCAACCTCAATGTCTATCTTGGCCAAAGTATGGTCTGGACGGNAAGAAAGACGCTCGGCA 389
Db |||||||
QY 363 GGTGACCAAGCACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCAGGCTGGA 422
Db |||||||
QY 390 GGTGACCAAGCACTCCCTCAGTGGCACCCAGGAAGACGGGCTGGATGACCCAGGCTGGA 449
Db |||||||
QY 423 AAAGCTGTGGCA CAAGGCGAAGACCTCTGGGAAATTTCTCCGGCGAAGAACTGGACAAGCT 482
Db |||||||
QY 450 AAAGCTGTGGCA CAAGGCGAAGACCTCTGGGAAATTTCTCCGGCGAAGAACTGGACAAGCT 509
Db |||||||
QY 483 CTGGCGGGAGTTCTTCGTCATCAAAAGAGAAAGTTACAGGATACAAAGTCTCTGCTGGAGAC 542
Db |||||||
QY 510 CTGGCGGGAGTTCTTCGTCATCAAAAGAGAAAGTTACAGGATACAAAGTCTCTGCTGGAGAC 569
Db |||||||
QY 543 CTTGAGCAGGACCGAAGAAATCCACGAGAAACGTCATTAGCCCTCGGAACCTGAGCGACAT 602
Db |||||||
QY 570 CTTGAGCAGGACCGAAGAAATCCACGAGAAACGTCATTAGCCCTCGGAACCTGAGCGACAT 629
Db |||||||
QY 603 CAAAGGCGAGCGTCTTCGACAGAGGCA CACGAGCTGAAAGGAGAACTGCGCAGCATCAA 662
Db |||||||
QY 630 CAAAGGCGAGCGTCTTCGACAGAGGCA CACGAGCTGAAAGGAGAACTGCGCAGCATCAA 689
Db |||||||
QY 663 CCAGGCGCTTGGACCGGCTCGGACAGGCTCAGCCACAGGGCTACAGCACTGAGGCTGAGTT 722
Db |||||||
QY 690 CCAGGCGCTTGGACCGGCTCGGACAGGCTCAGCCACAGGGCTACAGCACTGAGGCTGAGTT 749
Db |||||||
QY 723 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTTGGCGCAGTCCGCCAACCTCACGAGCAA 782
Db |||||||
QY 750 CGAGGAGCCAGGGTGATTGACCTGTGGGACCTTGGCGCAGTCCGCCAACCTCACGAGCAA 809
Db |||||||
QY 783 GGAGCTGGAGGGGTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAAGACACAA 842
Db |||||||
QY 810 GGAGCTGGAGGGGTTCCGGGAGGAGCTCAAGCACTTCGAAGCCAAATCGAAGACACAA 869
Db |||||||
QY 843 CCACTACCAAGACGAGCTGGAGNTTCCGCAAGAGCTGAGGCAAGGAGGAGGCTGGG 902
Db |||||||
QY 870 CCACTACCAAGACGAGCTGGAGNTTCCGCAAGAGCTGAGGCAAGGAGGAGGCTGGG 929
Db |||||||
QY 903 CGACGCGAGCGTGTGAGCCGAGCGCGAGAGACACGCCCTGCTGGAGGGCGGACCAA 962
Db |||||||
QY 930 CGACGCGAGCGTGTGAGCCGAGCGCGAGAGACACGCCCTGCTGGAGGGCGGACCAA 989
Db |||||||
QY 963 GGAGCTGGGCTACACGGTGAAGAGCATCTGAGGACCTGTCCGGCAGGATCTCCAGAGC 1022
Db |||||||
QY 990 GGAGCTGGGCTACACGGTGAAGAGCATCTGAGGACCTGTCCGGCAGGATCTCCAGAGC 1049
Db |||||||
QY 1023 TCGCGCCGAGGCA 1035
Db |||||
QY 1050 TCGGCAACGAA 1062
Db |||||
```

Search completed: April 4, 2005, 18:57:30
Job time : 10494 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2005, 16:02:35 ; Search time 132 Seconds
(without alignments)
3598.044 Million cell updates/sec

Title: US-10-600-862A-7
Perfect score: 6522
Sequence: 1 MRGPSALWLLALRTVLGS.....KVLDCVSLLMGEQFLVSWC 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:.*
1: Geneseq19808:.*
2: Geneseq19908:.*
3: Geneseq20008:.*
4: Geneseq20018:.*
5: Geneseq20028:.*
6: Geneseq20038a:.*
7: Geneseq20038b:.*
8: Geneseq20048:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4744.5	72.7	952	8 ADQ91896	Adq91896 Human glu
2	4744.5	72.7	953	7 ADD45733	Add45733 Human Pro
3	4741	72.7	883	7 ABM79002	Abm79002 Human alp
4	4724.5	72.4	950	8 ADF47521	Adf47521 Human GAA
5	4283.5	65.7	878	8 ABM84315	Abm84315 Human dia
6	4260.5	65.3	904	8 ABM84314	Abm84314 Human dia
7	3974	60.9	953	5 ABBS7174	Abbs7174 Mouse isc
8	3763.5	57.7	788	8 ABM84316	Abm84316 Human dia
9	2024	31.0	707	7 ADJ83134	Adj83134 Glucosyl
10	2007.5	30.8	638	8 ABM84317	Abm84317 Human dia
11	1900.5	29.1	1857	8 ADQ19560	Adq19560 Human PRO
12	1877.5	28.8	1827	2 AAU74090	Aau74090 Human hsi
13	1877.5	28.8	1827	4 AAU09028	Aau09028 Human suc
14	1877.5	28.8	1829	4 ABG14994	Abg14994 Novel hum
15	1660.5	25.5	357	4 AAB48844	Aab48844 Human RAP
16	1660.5	25.5	357	5 AAQ18621	Aaq18621 Human rec
17	1660.5	25.5	357	7 ADD44975	Add44975 Human Pro
18	1660.5	25.5	357	7 ADD44979	Add44979 Human Pro
19	1660.5	25.5	357	7 ADD44971	Add44971 Human Pro
20	1660.5	25.5	357	7 ADD44967	Add44967 Human Pro
21	1660.5	25.5	357	7 ADD44963	Add44963 Human Pro
22	1660.5	25.5	357	7 ADD44959	Add44959 Human Pro
23	1660.5	25.5	357	7 ADD46432	Add46432 Human Pro
24	1660.5	25.5	357	8 ADU75361	Adj75361 Marker ge
25	1660.5	25.5	357	8 ADQ91471	Aadj91471 Amino aci

RESULT 1

ADQ91896					
ID	ADQ91896	standard; protein; 952 AA.			
XX	AC	ADQ91896;			
XX	DT	21-OCT-2004 (first entry)			
XX	DE	Human glucosidase alpha acid (GAA) protein SEQ ID NO:2.			
XX	KW	chimeric protein; secretory signal; lysosomal protein;			
KW	KW	lysosomal acid alpha-glucosidase; glucosidase alpha acid; GAA; enzyme;			
KW	KW	lysosomal acid alpha-glucosidase deficiency; GAA deficiency;			
KW	KW	glycogen storage disease type II; GSD II; human; chromosome 17.			
XX	OS	Homo sapiens.			
XX	XX	WO2004064750-A2.			
XX	PD	05-AUG-2004.			
XX	PF	21-JAN-2004; 2004WO-US001453.			
XX	XX	22-JAN-2003; 2003US-0441789P.			
XX	XX	(UYDU-) UNIV DUKE.			
XX	PI	Koeberl DD, Sun B;			
XX	XX	WPI; 2004-571599/55.			
DR	DR	N-PSDB; ADQ91895, ADQ91897.			
DR	DR	GENBANK; NM_000152.			
XX	XX	New nucleic acid expressing lysosomal acid alpha-glucosidase (GAA)			
PT	PT	polypeptide, useful for preparing a composition for treating e.g.,			
PT	PT	glycogen storage disease type II.			
XX	XX	Example 13; SEQ ID NO 2; 128pp; English.			
XX	CC	The present invention describes an isolated nucleic acid (I) encoding a			
CC	CC	chimeric polypeptide comprising a secretory signal sequence operably			
CC	CC	linked to a lysosomal polypeptide. Also described: (1) a vector			
CC	CC	comprising the isolated nucleic acid; (2) a pharmaceutical formulation			
CC	CC	comprising the isolated nucleic acid in a carrier; (3) a cell comprising			
CC	CC	the isolated nucleic acid; (4) a chimeric polypeptide comprising a			
CC	CC	secretory signal sequence operably linked to a lysosomal polypeptide; (5)			
CC	CC	delivering a nucleic acid encoding a lysosomal polypeptide (preferably a			
CC	CC	lysosomal acid alpha-glucosidase (GAA)) to a cell; (6) producing a GAA			

ALIGNMENTS

26	1660.5	25.5	357	8	ADP23135	Adp23135 PRO polyp
27	1660.5	25.5	682	8	ADQ39604	Adq39604 Human myo
28	1660.5	25.5	682	8	ADQ39605	Adq39605 Human tis
29	1650	25.3	323	2	AAR56282	Aar56282 Human tis
30	1647	25.3	323	2	AAR80944	Aar80944 Receptor -
31	1498.5	23.0	902	6	ABU63302	Abu63302 Wild-type
32	1472	22.6	901	5	ABB09151	Abb09151 Buckwheat
33	1472	22.6	901	5	AAU97731	Aau97731 Common bu
34	1459.5	22.4	646	8	ADQ67734	Adq67734 Novel hum
35	1459.5	22.4	903	6	ABU63301	Abu63301 Wild-type
36	1457	22.3	901	5	ABB09152	Abb09152 Buckwheat
37	1457	22.3	901	5	AAU97732	Aau97732 Common bu
38	1403.5	21.5	877	2	AAW59040	Aaw59040 Barley al
39	1401.5	21.5	914	6	ABU63150	Abu63150 Wild-type
40	1392.5	21.4	969	7	ADC51564	Adc51564 Schizosac
41	1392.5	21.4	969	8	ADH61320	Adh61320 Schizosac
42	1387	21.3	877	6	ABU63305	Abu63305 Barley al
43	1381	21.2	877	6	ABU63304	Abu63304 Barley al
44	1380	21.2	877	6	ABU63308	Abu63308 Barley al
45	1380	21.2	877	6	ABU63307	Abu63307 Barley al

polypeptide in a cultured cell; (7) treating a deficiency of a lysosomal polypeptide or lysosomal acid alpha-glucosidase (GAA) in a subject; and (8) an isolated nucleic acid (ii) encoding a GAA polypeptide comprising: (a) a coding region encoding a GAA polypeptide; and (b) a 3' untranslated region where: (i) the 3' untranslated region (UTR) comprises a GAA 3' UTR comprising a deletion of at least 25 consecutive nucleotides, so that upon introduction into a cell, GAA polypeptide is produced at a higher level from the nucleic acid as compared with its production from a nucleic acid comprising a full-length GAA 3' UTR; or (ii) the 3' UTR is less than 200 nucleotides in length and comprises a segment that is heterologous to the GAA coding region, so that GAA polypeptide is overexpressed on introduction of the nucleic acid into a cell. (i) has hepatotropic activity, and can be used in gene therapy. The nucleic acid (i) encoding a chimeric polypeptide comprising a secretory signal sequence operably linked to a lysosomal acid alpha-glucosidase (GAA) polypeptide is useful for preparing a composition for treating lysosomal acid alpha-glucosidase (GAA) deficiency e.g., glycogen storage disease type II (GSD II). The present sequence represents the human GAA amino acid sequence, which is used in the exemplification of the present invention. The human GAA gene is located on chromosome 17, more specifically to 17q25.2-q25.3.

XX Sequence 952 AA;

Query Match 72.7%; Score 4744.5; DB 8; Length 952;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY	313	LEGRTELGVTVKHHQLDLSGRISRAAEATGAHGRPRAVPTQCDVPNSRFDCAPOK	372
DB	42	LSGSSPVLEETHPAHQGASRGPFR-----DAQAHFGRPRAVPTQCDVPNSRFDCAPOK	96
QY	373	AITQEOCEARGCCVPAKQGLQAGQMPGPPPSVSKLENLSSEMGVYATLITRTT	432
DB	97	AITQEOCEARGCCVPAKQGLQAGQMPGPPPSVSKLENLSSEMGVYATLITRTT	156
QY	433	PTFFPKDILTRLDVMMETENRHLFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFESE	492
DB	157	PTFFPKDILTRLDVMMETENRHLFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFESE	216
QY	493	PGVIVHRQLDGRVLLNTTVAPLFFADQFLOLSTSLPSQYITGLAEHLSPMLSTSWTRI	552
DB	217	PGVIVHRQLDGRVLLNTTVAPLFFADQFLOLSTSLPSQYITGLAEHLSPMLSTSWTRI	276
QY	553	TLNWRDLAPTPGANLVGSHPFYALBDGSAHGVFLNLSNAMDVLQPSALSWRSTGGI	612
DB	277	TLNWRDLAPTPGANLVGSHPFYALBDGSAHGVFLNLSNAMDVLQPSALSWRSTGGI	336
QY	613	LDVYIFLGPPEKSVVQOYLDVVGYPFMPYPWGLGFLCRWGYSSTAITRQVVENMTRAHF	672
DB	337	LDVYIFLGPPEKSVVQOYLDVVGYPFMPYPWGLGFLCRWGYSSTAITRQVVENMTRAHF	396
QY	673	PLDVQWNLDMYDSRDFTFNKDGPRDFPAMVQELHGGRRYMMIVDPAISSGGPAGSYR	732
DB	397	PLDVQWNLDMYDSRDFTFNKDGPRDFPAMVQELHGGRRYMMIVDPAISSGGPAGSYR	456
QY	733	PYDEGLRRGVFIINERGQPLIGKVPSTAFDPFTPTALAWEDMVAFHDOQVPPDGLW	792
DB	457	PYDEGLRRGVFIINERGQPLIGKVPSTAFDPFTPTALAWEDMVAFHDOQVPPDGLW	516
QY	793	IDNNEFSNFIIRGSDGCPNNELENPPYVGVGGTTLQAATICASSHQFLSTHYNLNLYG	852
DB	517	IDNNEFSNFIIRGSDGCPNNELENPPYVGVGGTTLQAATICASSHQFLSTHYNLNLYG	576
QY	853	LTEAIAASHRALVKARGTRPVIISRTFAGHGRVAGHTGDVWSSWQLASSVPEILQFNL	912
DB	577	LTEAIAASHRALVKARGTRPVIISRTFAGHGRVAGHTGDVWSSWQLASSVPEILQFNL	636
QY	913	LGVPPLGADVCGFLGNTSELCVRYTQLGAFYFPMENHNSLLSLPOEPYSFSEPAQAMR	972
DB	637	LGVPPLGADVCGFLGNTSELCVRYTQLGAFYFPMENHNSLLSLPOEPYSFSEPAQAMR	696
QY	973	KALTTRYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPVL	1032

DB	697	KALTTRYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPVL	756
QY	1033	QAGKAERTGYFPGLGTWYDLQTVPIEALGSLRPPRAAPREPAIHSEGWVTLPAPLDTINV	1092
DB	757	QAGKAERTGYFPGLGTWYDLQTVPIEALGSLRPPRAAPREPAIHSEGWVTLPAPLDTINV	816
QY	1093	HLRAGYIIPLOQPGGLTTTSTESRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTV	1152
DB	817	HLRAGYIIPLOQPGGLTTTSTESRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTV	876
QY	1153	IFLARNTIYNELVRVTSEGAGLQLOKVTLVGVATAPQOVLNSGVPSVFTTSPDTKVLD	1212
DB	877	IFLARNTIYNELVRVTSEGAGLQLOKVTLVGVATAPQOVLNSGVPSVFTTSPDTKVLD	936
QY	1213	ICVSLLMGESQFLVSWC 1228	
DB	937	ICVSLLMGESQFLVSWC 952	
RESULT 2			
ADD45733			
ID	ADD45733 standard; protein; 953 AA.		
XX	AC ADD45733;		
XX	AC ADD45733;		
DT	29-JAN-2004 (first entry)		
XX	Human Protein Y00839, SEQ ID NO 11401.		
XX	Human; pain; neuronal tissue; gene therapy;		
KW	spinal segmental nerve injury; chronic constriction injury; CCI;		
KW	spared nerve injury; SNI; Chung.		
XX	Homo sapiens.		
XX	WO2003016475-A2.		
XX	27-FEB-2003.		
XX	14-AUG-2002; 2002WO-US025765.		
XX	14-AUG-2001; 2001US-0312147P.		
PR	01-NOV-2001; 2001US-0346382P.		
PR	26-NOV-2001; 2001US-0333347P.		
XX	(GEHO) GEN HOSPITAL CORP.		
PA	(FARB) BAYER AG.		
XX	Woolf C, D'urso D, Befort K, Costigan M;		
PI	WPI: 2003-268312/26.		
XX	GENBANK; Y00839.		
DR	New composition comprising two or more isolated polypeptides, useful for		
XX	preparing a medicament for treating pain in an animal.		
PS	Claim 1; Page; 1017pp; English.		
XX	The invention discloses a composition comprising two or more isolated rat		
CC	or human polynucleotides or a polynucleotide which represents a fragment,		
CC	derivative or allelic variation of the nucleic acid sequence. Also		
CC	claimed are a vector comprising the novel polynucleotide, a host cell		
CC	comprising the vector, a method for identifying a nucleotide sequence		
CC	which is differentially regulated in an animal subjected to pain and a		
CC	kit to perform the method, an array, a method for identifying an agent		
CC	that increases or decreases the expression of the polynucleotide sequence		
CC	that is differentially expressed in neuronal tissue of a first animal		
CC	subjected to pain, a method for identifying a compound which regulates		
CC	the expression of a polynucleotide sequence which is differentially		
CC	expressed in an animal subjected to pain, a method for identifying a		
CC	polynucleotides, a method for producing a pharmaceutical composition, a		

method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification), a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 953 AA;

Query Match 72.7%; Score 4744.5; DB 7; Length 953;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY 313 LEGRTKELGYVVKHQLDLSGRISRAAEATGAHPRPRAVPTQCDVPPNSRFDCAPOK 372
DB 42 LSGSPVLEETHAHOQASRPGR-----DAQAHGPRPRAVPTQCDVPPNSRFDCAPOK 96
QY 373 AITQEQCEARGCCYIPAKQLOQAQOMPCCPPSPSYKLENLSSEMGTATLTRT 432
DB 97 AITQEQCEARGCCYIPAKQLOQAQOMPCCPPSPSYKLENLSSEMGTATLTRT 156
QY 433 PTFPPKDIILTRLDVMMETENRHLFTIKDPANRRYEVPLETPRVHGRAPSPLYSVEFSEE 492
DB 157 PTFPPKDIILTRLDVMMETENRHLFTIKDPANRRYEVPLETPRVHGRAPSPLYSVEFSEE 216
QY 493 PFGVIVHRQLDGRVLLNTTVAFLFFADQLQLSTLPSQYITGLAEHLSPLMLSTWTRI 552
DB 217 PFGVIVHRQLDGRVLLNTTVAFLFFADQLQLSTLPSQYITGLAEHLSPLMLSTWTRI 276
QY 553 TLNWRDLAPTGANLYGSHPPYALAEDEGSAHGCVLLNSNAMDVVLQSPALSWRSTGGI 612
DB 277 TLNWRDLAPTGANLYGSHPPYALAEDEGSAHGCVLLNSNAMDVVLQSPALSWRSTGGI 336
QY 613 LDVYIFLGPPEKSVQOYLDDVGYPPMPYPVWGLGFLCRWGSSTAITQVVENMTRAHF 672
DB 337 LDVYIFLGPPEKSVQOYLDDVGYPPMPYPVWGLGFLCRWGSSTAITQVVENMTRAHF 396
QY 673 PLDVQWNLDDYMDSRDFTFNKDGFRDFPAMVQELHQGGRYMMIVDPAISSGGPAGSYR 732
DB 397 PLDVQWNLDDYMDSRDFTFNKDGFRDFPAMVQELHQGGRYMMIVDPAISSGGPAGSYR 456
QY 733 PYDEGLRRGVFTNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDMVAFHDOQVPPDGLW 792
DB 457 PYDEGLRRGVFTNETGQPLIGKWPFGSTAFDPFTNPTALAWMEDMVAFHDOQVPPDGLW 516
QY 793 IDWNEPSNFRGSEDCPNNELENPPYVGVWGGTLQAATICASSHQPSTHYNLNLXG 852
DB 517 IDWNEPSNFRGSEDCPNNELENPPYVGVWGGTLQAATICASSHQPSTHYNLNLXG 576
QY 853 LTEATASHRALVKARTRFPVISRSTFAGHGRVAGHTGDMVSSWQLASSVPEILQFNL 912
DB 577 LTEATASHRALVKARTRFPVISRSTFAGHGRVAGHTGDMVSSWQLASSVPEILQFNL 636
QY 913 LGVPLVGADVCGFLGNTSBEICVRWTLQGAFFPMMNHNLSLPOEPYSFSEPAQOAMR 972
DB 637 LGVPLVGADVCGFLGNTSBEICVRWTLQGAFFPMMNHNLSLPOEPYSFSEPAQOAMR 696
QY 973 KALTLYALLPHLYTLFQAHVAGETVARPLFLFPPKDSSTWTDHQLLWGLALLTPVL 1032
DB 697 KALTLYALLPHLYTLFQAHVAGETVARPLFLFPPKDSSTWTDHQLLWGLALLTPVL 756
QY 1033 QAGKAEVTVGFLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEQWVTLPAPLDTINV 1092
DB 757 QAGKAEVTVGFLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEQWVTLPAPLDTINV 816
QY 1093 HLRAGYIIPLOQPGGLTTTTSROQPMALVALTKGSGEARGELFWDDGESLEVLERGAYTQV 1152

DB 817 HLRAGYIIPLOQPGGLTTTTSROQPMALVALTKGSGEARGELFWDDGESLEVLERGAYTQV 876
QY 1153 IFLARNNTIVNELVRVTSEAGLQKQVTLGVATAPQOQVLSNGVPSVSNFTYSPDTKVLD 1212
DB 877 IFLARNNTIVNELVRVTSEAGLQKQVTLGVATAPQOQVLSNGVPSVSNFTYSPDTKVLD 936
QY 1213 ICVSLLMGEQFLVSWC 1228
DB 937 ICVSLLMGEQFLVSWC 952

RESULT 3

ABM79002
ID ABM79002 standard; protein; 883 AA.

XX AC ABM79002;
XX DT 15-JAN-2004 (first entry)

XX Human alpha-galactosidase (mature polypeptide).

XX Human; alpha-glucosidase; lysosome; enzyme; Pompe disease;
transgenic plant.

XX Homo sapiens.

XX WO2003073839-A2.

XX 12-SEP-2003.

XX 03-MAR-2003; 2003WO-IT000120.

XX 01-MAR-2002; 2002IT-RM000115.

XX (PLAN-) PLANTECHNO SRL.

XX Fogher C, Reggi S;

XX WPI; 2003-712829/67.

XX N-PSDB; ACF80584.

XX New genetically transformed plant that can produce a lysosomal enzyme of animal or human origin, useful for preparing a medicament for enzyme replacement therapy in Gaucher, Anderson-Fabry or Pompe disease.

XX Example 10; Page 51-53; 53pp; English.

XX The present sequence is the protein sequence of human alpha-glucosidase (GAA). A deficiency of this enzyme causes Pompe disease. The invention is based on the discovery that lysosomal enzymes, such as GAA, can be expressed in seed storage organs in a form which is stable (over 12 months in stored seeds) enzymatically active and in a high amount suitable for medical use. An expression cassette includes the GAA coding sequence (minus the native signal sequence), and basic 7S soy globulin promoter and signal sequences. Such constructs can be used to express human GAA in transgenic plants, especially plants having a high protein content, e.g. legumes, cereals and tobacco

XX Sequence 883 AA;

Query Match 72.7%; Score 4741; DB 7; Length 883;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 882; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 346 AHGPRPRAVPTQCDVPPNSRFDCAPOKAITQEQCEARGCCYIPAKQLOQAQOMPCCFF 405

DB 1 AHGPRPRAVPTQCDVPPNSRFDCAPOKAITQEQCEARGCCYIPAKQLOQAQOMPCCFF 60

QY 406 PPSYPSYKLENLSSEMGTATLTRTFFPKDILTRLDVMMETENRHLFTIKDPANR 465

DB 61 PPSYPSYKLENLSSEMGTATLTRTFFPKDILTRLDVMMETENRHLFTIKDPANR 120

QY 466 RYEVPLETRVRSRAPSPLYSVEFSEPRGVIHROLDORVLLNTTVAELFADQFLQLS 525
DB 121 RYEVPLETRVRSRAPSPLYSVEFSEPRGVIHROLDGRVLLNTTVAELFADQFLQLS 180
QY 526 TSLPSQYITGLAEHLSPMLSTSWTRITLWNRLAFTPGANLYGSHPPFYLALEDGGSAGH 585
DB 181 TSLPSQYITGLAEHLSPMLSTSWTRITLWNRLAFTPGANLYGSHPPFYLALEDGGSAGH 240
QY 586 VFLLSNAMDMVVLQSPALSWRSTGILDVYIFLGPPEKSVVQQYLDVVGVYFMPYPYMG 645
DB 241 VFLLSNAMDMVVLQSPALSWRSTGILDVYIFLGPPEKSVVQQYLDVVGVYFMPYPYMG 300
QY 646 GFHLRWGYSSTAITRQVENMTRAFPLDVQWDLDMDSRRDFTFNKDGFRDPFAMVQ 705
DB 301 GFHLRWGYSSTAITRQVENMTRAFPLDVQWDLDMDSRRDFTFNKDGFRDPFAMVQ 360
QY 706 ELHQGRRYMMIVDPAISSGPGAGSRPYDEGLRRCGVFTNETGOLIKGVWPGSTAPDP 765
DB 361 ELHQGRRYMMIVDPAISSGPGAGSRPYDEGLRRCGVFTNETGOLIKGVWPGSTAPDP 420
QY 766 FNTPTALAWMEDVAFHFQVDPEDGILDMWNEPNSFIRGSDGCPNNELENPPYVGVVG 825
DB 421 FNTPTALAWMEDVAFHFQVDPEDGILDMWNEPNSFIRGSDGCPNNELENPPYVGVVG 480
QY 826 GTLQAAATICASSHQFLSTHYNLHNLVGLTEATASHRALVKARTRPFVSRSTFAGHG 885
DB 481 GTLQAAATICASSHQFLSTHYNLHNLVGLTEATASHRALVKARTRPFVSRSTFAGHG 540
QY 886 AGHWTGDVWSSWQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRTQLGAF 945
DB 541 AGHWTGDVWSSWQLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRTQLGAF 600
QY 946 FMRNHNLSLLPOEYFSFSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPL 1005
DB 601 FMRNHNLSLLPOEYFSFSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPL 660
QY 1006 EPPKDSSTWTDHQLWGBALLITPVLOAKGAEVTGYFPLGTWYDILQTVPIBALGSL 1065
DB 661 EPPKDSSTWTDHQLWGBALLITPVLOAKGAEVTGYFPLGTWYDILQTVPIBALGSL 720
QY 1066 PAAPRPALHSEGWTLPAFLDTINVHLRAGVILPQGPGLTTTSSROQPWALAVLTK 1125
DB 721 PAAPRPALHSEGWTLPAFLDTINVHLRAGVILPQGPGLTTTSSROQPWALAVLTK 780
QY 1126 GGEARGELFWDDGESLEVLERGAYTOVIFLARNTIVNELVRVTSAGLQKQVTLGV 1185
DB 781 GGEARGELFWDDGESLEVLERGAYTOVIFLARNTIVNELVRVTSAGLQKQVTLGV 840
QY 1186 ATAPQOVLNGVPVSNFTYSPDTKVLDICVSLLMGQFLVSWC 1228
DB 841 ATAPQOVLNGVPVSNFTYSPDTKVLDICVSLLMGQFLVSWC 883

RESULT 4

ID ADF47521 standard; protein; 950 AA.

XX ADF47521;

AC ADF47521;

XX ADF47521;

DT 26-FEB-2004 (first entry)

XX Human GAA protein SEQ ID NO:24.

DE Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

XX Human GAA protein SEQ ID NO:24.

KW infantile neuronal ceroid lipofuscinosis; Ehlers-Danlos syndrome type VI;
KW congenital glycosylation disorder.
OS Homo sapiens.
PN WO2003102583-A1.
XX 11-DEC-2003.
PD 29-MAY-2003; 2003WO-US017211.
XX 29-MAY-2003; 2003US-0384452P.
PF 05-JUN-2002; 2002US-0384452P.
XX 06-SEP-2002; 2002US-0408816P.
PR 16-OCT-2002; 2002US-00272531.
XX 06-FEB-2003; 2003US-0445734P.
XX (SYMB-) SYMBIONTICS INC.

PI Lebowitz JH, Beverley SM, Sly WS;
XX WPI; 2004-035473/03.
XX N-PSDB; ADF47520.
XX Novel underglycosylated targeted therapeutic comprising therapeutic agent
PT active in human lysosome, lysosomal targeting domain binds to human
PT cation-independent mannose-6-phosphate receptor, for treating Pompe
PT disease.
XX Example 13B; SEQ ID NO 24; 137pp; English.

XX The present invention describes an underglycosylated targeted therapeutic
CC (GT) which comprises: (a) a therapeutic agent that is therapeutically
CC active in human lysosome; and (b) a lysosomal targeting domain that binds
CC an extracellular domain of human cation-independent mannose-6-phosphate
CC receptor (CM) and (i) does not bind a mutin in which amino acid 1572 of
CC CM is changed from isoleucine to threonine; and (ii) binds the mutin
CC with dissociation constant at least ten times the dissociation constant
CC for binding CM. Also described: (1) an underglycosylated therapeutic
CC fusion protein (FP) comprising a therapeutic domain and a subcellular
CC targeting domain that binds to an extracellular domain of a receptor on
CC an exterior surface of a cell, and upon internalisation of the receptor,
CC permits localisation of the therapeutic domain to a subcellular
CC compartment where the therapeutic domain is therapeutically active; and
CC (2) production of GT. GT has nephrotropic activity, and can be used in
CC enzyme replacement therapy, cell therapy and gene therapy. The FP can be
CC used for treating a lysosomal storage disease patient by administering FP
CC to the patient. GT can be used for treating a patient by identifying a
CC targeting moiety that binds CM in a mannose-6-phosphate independent
CC manner, synthesising GT comprising a therapeutic agent that is
CC therapeutically active in a mammalian lysosome and a targeting moiety
CC that binds CM in a mannose-6-phosphate independent manner and
CC administering GT to the patient, where the targeting moiety is identified
CC by screening a nucleic acid or peptide library. GT is useful for treating
CC metabolic disease, lysosomal storage diseases and associated enzyme
CC defects such as Pompe disease, Tay-Sachs disease, Sandhoff disease, Fabry
CC disease, Gaucher disease, Krabbe disease, Wolman disease, Hurler
CC syndrome, Hunter syndrome, Sly syndrome, Schindler disease, infantile
CC sialic acid storage disease, Batten disease, infantile neuronal ceroid
CC lipofuscinosis, Ehlers-Danlos syndrome type VI and congenital disorders
CC of glycosylation. The present sequence is used in the exemplification of
CC the present invention.

XX Sequence 950 AA;
SQ Query Match 72.4%; Score 4724.5; DB 8; Length 950;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 888; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY 313 LEGRTELKELGYTVKHLQDLISRAAEATGAGHPRPRAVPTQCDVPPNSRFDCA 372
DB 42 LSGSPVLEETHPAHQOQNSRGPGR-----DAQAHGPRPRAVPTQCDVPPNSRFDCA 96

```
QY 373 AITQCEARGCCYIPAKQGLQAGQWPCWCFPPSPSYKLENLSSEMGGYATLTRTT 432
DB 97 AITQCEARGCCYIPAKQGLQAGQWPCWCFPPSPSYKLENLSSEMGGYATLTRTT 156
QY 433 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHGRAPSPLYSVFSEE 492
DB 157 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHGRAPSPLYSVFSEE 216
QY 493 PFGVIVRHQDGRVLLNTTVAFLFFADQFLQSLSTPSQYITGLAEHLSPMLSTWTRI 552
DB 217 PFGVIVRHQDGRVLLNTTVAFLFFADQFLQSLSTPSQYITGLAEHLSPMLSTWTRI 276
QY 553 TLNWRDLAPTPGANLYGSHFYLALDGGSAHGVLNLSNAMDVVLQSPALSWRSTGGI 612
DB 277 TLNWRDLAPTPGANLYGSHFYLALDGGSAHGVLNLSNAMDVVLQSPALSWRSTGGI 336
QY 613 LDVYIFLGPPKSVVQOYLQDVGVPPMPYVGLGHLCRWGYSTALTROVVNMTRAHP 672
DB 337 LDVYIFLGPPKSVVQOYLQDVGVPPMPYVGLGHLCRWGYSTALTROVVNMTRAHP 396
QY 673 PLDVQWMDLDYMSRRDFTFNKDGFRDPAMVQELHOGGRYVMIVDPAISSSGPAGSYR 732
DB 397 PLDVQWMDLDYMSRRDFTFNKDGFRDPAMVQELHOGGRYVMIVDPAISSSGPAGSYR 456
QY 733 PYDEGLRRGVITNETQPLIGKWPGSTAPPDFTNPTALAWEDMVAEFHDQVDFGLW 792
DB 457 PYDEGLRRGVITNETQPLIGKWPGSTAPPDFTNPTALAWEDMVAEFHDQVDFGLW 516
QY 793 IDNNEPSNFRGSEDCPNNELENPPYVGVGGTLOAATICASSHQFLSTHYNLNLYG 852
DB 517 IDNNEPSNFRGSEDCPNNELENPPYVGVGGTLOAATICASSHQFLSTHYNLNLYG 576
QY 853 LTBALASHRALVKARGTRPFIISRSFAGHGRVAGHTGDMVSWEQGLASSVPEILQFNL 912
DB 577 LTBALASHRALVKARGTRPFIISRSFAGHGRVAGHTGDMVSWEQGLASSVPEILQFNL 636
QY 913 LGVPLVGADVCGFLGNTSEELCVRWITQLGAFYFPMRNHNSLLSLPQEPYFSFSPAQAMR 972
DB 637 LGVPLVGADVCGFLGNTSEELCVRWITQLGAFYFPMRNHNSLLSLPQEPYFSFSPAQAMR 696
QY 973 KALTLYALLPHLYTLFPHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVL 1032
DB 697 KALTLYALLPHLYTLFPHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVL 756
QY 1033 QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAPAPREPAIHSEGOVTLPAPLDTINV 1092
DB 757 QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAPAPREPAIHSEGOVTLPAPLDTINV 816
QY 1093 HLRAGYIIPQGPGLTTESRQQPMALAVALTGGEARGELFWDDGESLEVLERGAYTQV 1152
DB 817 HLRAGYIIPQGPGLTTESRQQPMALAVALTGGEARGELFWDDGESLEVLERGAYTQV 876
QY 1153 IFLARNTIIVELVRVTSEAGLQLOKVTVLGVATAPQOVLNGVPVSNFTYSPDTRKVL 1212
DB 877 IFLARNTIIVELVRVTSEAGLQLOKVTVLGVATAPQOVLNGVPVSNFTYSPDTRKVL 936
QY 1213 ICVSLLMGEQFLVS 1226
DB 937 ICVSLLMGEQFLVS 950
RESULT 5
ABM84315
ID ABM84315 standard; protein; 878 AA.
XX AC ABM84315;
XX DT
XX 18-NOV-2004 (first entry)
DE Human diagnostic and therapeutic pprotein SEQ ID NO:4564.
XX KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
```

```
OS Homo sapiens.
PN WO2004023973-A2.
XX 25-MAR-2004.
PD 12-SEP-2003; 2003WO-US028227.
PF 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX (INCY-) INCYTE CORP.
PA Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Deleagane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kitson ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX WPI: 2004-329368/30.
DR N-PSDB; ACM42967.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 878 AA;
```

```
Query Match 65.7%; Score 4283.5; DB 8; Length 878;
Best Local Similarity 88.9%; Pred. No. 0;
Matches 814; Conservative 3; Mismatches 20; Indels 79; Gaps 2;

QY 313 LBGRTKELGYTVKKHLQDLISGRISARAEATGAHGRPRVPTQCDVPPNSRFDCAADK 372
DB 42 LSGSSPVLSETHPAHQOGASRPGPR-----DAQAHGRPRVPTQCDVPPNSRFDCAADK 96
QY 373 AITQCEARGCCYIPAKQGLQAGQWPCWCFPPSPSYKLENLSSEMGGYATLTRTT 432
DB 97 AITQCEARGCCYIPAKQGLQAGQWPCWCFPPSPSYKLENLSSEMGGYATLTRTT 156
QY 433 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHGRAPSPLYSVFSEE 492
DB 157 PTFPPKDLRLDVMETENRHLFTIKDPANRREYVPLETPRVHGRAPSPLYSVFSEE 216
QY 493 PFGVIVRHQDGRVLLNTTVAFLFFADQFLQSLSTPSQYITGLAEHLSPMLSTWTRI 552
DB 217 PFGVIVRHQDGRVLLNTTVAFLFFADQFLQSLSTPSQYITGLAEHLSPMLSTWTRI 276
QY 553 TLNWRDLAPTPGANLYGSHFYLALDGGSAHGVLNLSNAMDVVLQSPALSWRSTGGI 612
|||||
```

Db	277	TLWNRDLAPTGANLYGSHPPFYALBEDGSAHGCVFLNLSNAMDVLQSPALSWRSTGGI	336	PI	Harthshorne TA, Suchorolski MT, Alcus CM, Pitts SJ, Elder LV;
QY	613	LDVYIFLGPPEKSVVQQYLDVVGYPMPYPWGLGFLHLCRWGSSYTAITRQVVENMTRAHF	672	PI	Mooney EM, Deleage AM, Panesar IS, Banville SC, Reddy TP;
Db	337	LDVYIFLGPPEKSVVQQYLDVVGYPMPYPWGLGFLHLCRWGSSYTAITRQVVENMTRAHF	396	PI	Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
QY	673	PLDVQNDLDYMDSRDFTFNKDGFRDPPAMVQELHGGRRYMMIVDPAISSGPGAGSYR	732	PI	Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
Db	397	PLDVQNDLDYMDSRDFTFNKDGFRDPPAMVQELHGGRRYMMIVDPAISSGPGAGSYR	456	PI	Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kilton ES;
QY	733	PYDEGLRGGVFTNETGQPLIGKWPGSTAFDPFTNPTALAWMEDVAFHFDQVDPDGLM	792	PI	Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
Db	457	PYDEGLRGGVFTNETGQPLIGKWPGSTAFDPFTNPTALAWMEDVAFHFDQVDPDGLM	516	PI	Patury S, Shi X, Suarez CU;
QY	793	IDMNEPSNIRGSDCPCNNLENPYPVGVGGTLOAATICASSHQFLSTHYNLHLYG	852	XX	WP1; 2004-329368/30.
Db	517	IDMNEPSNIRGSDCPCNNLENPYPVGVGGTLOAATICASSHQFLSTHYNLHLYG	576	XX	N-PSDB; ACN42966.
QY	853	LTEAIAASHRALVKARTRPFVISRSTFAGHRYAGHWTGDVWSSWQLASSVPEILQFNL	912	XX	New diagnostic and therapeutic polynucleotides and polypeptides, useful
Db	577	LTEAIAASHRALVKARTRPFVISRSTFAGHRYAGHWTGDVWSSWQLASSVPEILQFNL	636	PT	in diagnosing a condition, disease or disorder associated with human
QY	913	LGVPVLGADVCGPLGNTSEELCVRWTLQCAFYPFEMNHNLSLSLPQEPYFSEPAQQAHR	972	PT	molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
Db	637	LGVPVLGADVCGPLGNTSEELCVRWTLQCAFYPFEMNHNLSLSLPQEPYFSEPAQQAHR	696	PT	in gene mapping.
QY	973	KALTLYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVL	1032	PS	Claim 27; Page; 190pp; English.
Db	697	KALTLYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVL	738	XX	The invention relates to novel diagnostic and therapeutic polynucleotides
QY	1033	QAGKAEVGVFLGTWYDLOQVPIEALGSLPPPPAAREPAIHSEQWTLPAFLDTINV	1092	CC	selected from one of the 2722 sequences defined in the specification. A
Db	739	-----	742	CC	polynucleotide of the invention may have a use in gene therapy. The human
QY	1093	HLRAGYIIPLOGPLTTTSROQPMALAVATKGGARGELFWDDGESLEVLERGAYTOV	1152	CC	diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
Db	743	HLRAGYIIPLOGPLTTTSROQPMALAVATKGGARGELFWDDGESLEVLERGAYTOV	802	CC	used to diagnose a particular condition, disease or disorder associated
QY	1153	IFLARNTTVNELVRVTSAGAGLQKVTVLGVATAPQOVLNGVPSNFTYSPDTKVLD	1212	CC	with human molecules, e.g. cell proliferative disorders,
Db	803	IFLARNTTVNELVRVTSAGAGLQKVTVLGVATAPQOVLNGVPSNFTYSPDTKVLD	862	CC	autoimmune/inflammatory disorder, developmental disorder, endocrine
QY	1213	ICVSLLMGQFLVSWC 1228		CC	disorder, neurological disorders, gastrointestinal disorders, or
Db	863	ICVSLLMGQFLVSWC 878		CC	infections caused by virus, bacteria, fungi or parasite. The dithp
RESULT 6				CC	molecules may also be used in genetic mapping, in identifying individuals
ABM84314				CC	from minute biological samples, in detecting single nucleotide
XX	ABM84314 standard; protein; 904 AA.			CC	polymorphisms, as molecular weight markers, and for somatic or germline
AC	ABM84314;			CC	gene therapy. The present sequence data for this patent is not represented in
XX	18-NOV-2004 (first entry)			CC	the printed specification, but was obtained in electronic format directly
DT	Human diagnostic and therapeutic pprotein SEQ ID NO:4563.			CC	from WIPO at www.wipo.int/pct/en/sequences/listing.htm
DE	gene therapy; human diagnostic and therapeutic polynucleotide; dithp.			XX	Sequence 904 AA;
XX	Homo sapiens.			QY	Query Match 65.3%; Score 4260.5; DB 8; Length 904;
XX	WO2004023973-A2.			Db	Best Local Similarity 86.4%; Pred. No. 0;
XX	25-MAR-2004.			QY	Matches 814; Conservative 3; Mismatches 20; Indels 105; Gaps 3;
PP	12-SEP-2003; 2003WO-US028227.			Db	313 LEGRTKELGYTVKHLQDLSGRISARAEATGAHGRPRAPVPTOCDVPPNSRFPDAPDK 372
XX	12-SEP-2002; 2002US-0410259P.			QY	42 LSGSPVLEETHFAHQOGASRGPGR-----DAQAHPGRPRAPVPTOCDVPPNSRFPDAPDK 96
PR	12-SEP-2002; 2002US-0410260P.			Db	373 AITQEQCEARGCCYIPAKQGLQGAQWQPWCFPPSPSYKLENLSSEMGTATLTRTT 432
XX	(INCY-) INCYTE CORP.			QY	97 AITQEQCEARGCCYIPAKQGLQGAQWQPWCFPPSPSYKLENLSSEMGTATLTRTT 156
XX	Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;			QY	433 PTFPPKDIILTLRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEE 492
PI				Db	157 PTFPPKDIILTLRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEE 216

QY 767 TNPATLAWMEDVVAEFDQVPPDGLWIDMNEPSNFIRGSDGCPNNELENPPYVPGVGG 826
 DB 517 TNPATLAWMEDVVAEFDQVPPDGLWIDMNEPSNFIRGSDGCPNNELENPPYVPGVGG 576
 QY 827 TLQATATICASSHQFLSTHYNLNLGLTALASHRALVKARGTRPPVISRSTFAGHGRYA 886
 DB 577 TLQATATICASSHQFLSTHYNLNLGLTALASHRALVKARGTRPPVISRSTFAGHGRYA 636
 QY 887 GHWTGDVWSSWEQCLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRTVQLGAFYPP 946
 DB 637 GHWTGDVWSSWEQCLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRTVQLGAFYPP 696
 QY 947 MRNHNLSLLPQEPYSPFPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLLE 1006
 DB 697 MRNHNLSLLPQEPYSPFPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLLE 756
 QY 1007 FPKDSSTMTVDHQLLWGEALLTPVLQAGKAEVTGYFPLGTWYDLQTVFIEALGSLPPPP 1066
 DB 757 FPKDSSTW----- 764
 QY 1067 AAPREPAIHSEGWTLPAPLDTINVHLRAGYIIPQGFGLTTTSSRQOPMALAVALTGK 1126
 DB 765 -----TINVHLRAGYIIPQGFGLTTTSSRQOPMALAVALTGK 802
 QY 1127 GEARGELFWDDGESLEVLRGAYTQVIFLARNTIINELVRVTSEGAGLOLOKVTVLGVA 1186
 DB 803 GEARGELFWDDGESLEVLRGAYTQVIFLARNTIINELVRVTSEGAGLOLOKVTVLGVA 862
 QY 1187 TAPQOVLNGVPSNFTYSPDTKVLDCVSLLMGEQFLVSWC 1228
 DB 863 TAPQOVLNGVPSNFTYSPDTKVLDCVSLLMGEQFLVSWC 904

RESULT 7

ABB57174
 ID ABB57174 standard; protein; 953 AA.
 XX
 AC ABB57174;
 DT 07-MAR-2002 (first entry)
 XX
 DE Mouse ischaemic condition related protein sequence SEQ ID NO:422.
 KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasoepaetic ischaemia; ischaemic condition; ischaemic disease.
 XX
 OS Mus musculus.
 XX
 PN WO200188188-A2.
 XX
 PD 22-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-JP004192.
 XX
 PR 18-MAY-2000; 2000JP-00145977.
 XX
 PA (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
 XX
 PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
 XX
 DR WPI; 2002-034733/04.
 DR N-PSDB; ABI99468.
 XX
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or by
 PT determining the expression profile of a gene group comprising these
 PT genes.
 PS Claim 2; Page 1154-1158; 2690pp; English.
 CC The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (i) in a test sample or determining the expression profile of a

CC gene group in the sample comprising genes selected from (i). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasoepaetic ischaemia) by measuring
 CC expression levels of particular genes (ABI99202 to ABI99912, encoding the
 CC protein sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The expression
 CC levels or expression profiles produced by these genes are used as an
 CC indicator when screening for ischaemic condition-improving drugs or
 CC therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR
 CC primers for a mouse ischaemic condition related sequence, which are used
 CC in the exemplification of the present invention

SQ Sequence 953 AA;

Query Match 60.9%; Score 3974; DB 5; Length 953;

Best Local Similarity 79.8%; Pred. No. 0;

Matches 739; Conservative 73; Mismatches 96; Indels 18; Gaps 4;

QY 320 LGYTVKKHL-----QDL-----SGRISRAAEATGAHPG-----RPRVPTQCDVPP 362
 DB 27 LGHLMRLMLLPQDLHSSGLWKTYRPHHOGYEPGLHTIQOTEQKPEAPTQCDVPP 86
 QY 363 NSRFCDAPDKAITQOCCEARGCCYIPAKQGLQAGQWQPFCFPSPSPSYKLENLSSEM 422
 DB 87 SRFCDAPDKGISQCCCEARGCCYIPAGQVLKEPQIGQPCWCFPPSPSYRLENLSSTES 146
 QY 423 GYTATLTRTPTFFPKDILTLRLDVMETENLHFTIKDPANRREYEVPLETPRVHSRAPS 482
 DB 147 GYTATLTRTPTFFPKDVLTLQLEVLMTDSLHFKIKDPASKRYEVEPLETPRVLSQAPS 206
 QY 483 PLYSVFSEEPFGVIVHQLDORVLLNTTVAPLFPADQFLQSLPSQYITGLAEHLSP 542
 DB 207 PLYSVFSEEPFGVIVRRKLGRLVLLNTTVAPLFPADQFLQSLPSLPAQHITGLGSHLSP 266
 QY 543 LMLSTSWTRITLWNRDLAPTGCANLYGSHPPFLALEDGGSAGHGVPLNLSNAMDVVLQSP 602
 DB 267 LMLSTDWARTILWNRDTPPSQTNLYGSHPPFLALEDGGLANGVFLNLSNAMDVVLQSP 326
 QY 603 ALSMRSTGGILDVYIFLQPEPKSVVQYLDVVGYPFMPYWGVLGPHLCRWGYSSTAIRQ 662
 DB 327 ALTWSTGGILDVYVFLGPEPKSVVQYLDVVGYPFMPYWGVLGPHLCRWGYSSTAIRQ 386
 QY 663 VVENNTRAHFPLDVQWNLDDYMSRRDFTFNKDGFRDPPAMVOELHOGGRRYMMIVDPAI 722
 DB 387 VVENNTRTHFPLDVQWNLDDYMDARDEFTFNQDSFADPDMDVRDVHQGGRRYMMIVDPAI 446
 QY 723 SSSGPAGSVRYPVDEGLRGVFTTNETGQPLICKVMPGSTAFDPDFTNPTALAWMEDVVAEF 782
 DB 447 SSAGPAGSVRYPVDEGLRGVFTTNETGQPLICKVMPGSTAFDPDFTNPTALAWMEDVVAEF 506
 QY 783 HDQVPFDGLWIDMNEPSNFIRGSDGCPNNELENPPYVPGVGGTLLQAAATICASSHQFLS 842
 DB 507 HQQVPFDGWLMDMNEPSNFVRGSGQCPNNELENPPYXGVVGGTLLQAAATICASSHQFLS 566
 QY 843 THYNLNLGLTEATASHRALVKARGTRPPVISRSTFAGHGRYAGHWTGDVWSSWEQLAS 902
 DB 567 THYNLNLGLTEATASHRALVKARGTRPPVISRSTFAGHGRYAGHWTGDVWSSWEQLAS 636
 QY 903 SVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRTVQLGAFYPPMRNHNLSLLPQEPYS 962
 DB 627 SVPEILQFNLLGVPLVGADVCGFIDTSEELCVRTVQLGAFYPPMRNHNLSLLPQEPYR 686
 QY 963 FSEPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTMTVDHQLLW 1022
 DB 687 FSEPAQAMRKALRYALLPYLYTLFRAHVRGDTVARPLFLFPREDSTWSVDRLWLW 746
 QY 1023 GBALLITPVLOAGKAEVTGYFPLGTWYDLQTVFIEALGSLPPPPAAR-EPAIHSEGWV 1081
 DB 747 GPALLITPVLEPGKTEVTGYFPKGTWYNNQVVSVDLSGLTLPSPSSASSFRSAVQSGQWL 806
 QY 1082 TLPAPLDTINVHLRAGYIIPQGFGLTTTSSRQOPMALAVALTGKGARGELFWDDGESL 1141
 DB 807 TLEAPLDTINVHLREGYIIPQGFGLTTTSSRQOPMALAVALTASGEADGELFWDDGESL 866

QY 1142 EVLERGAYTQVIFLARNNTIIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSN 1201
DB 867 AVLEHAYTLVFSAKNTIVNKLVRVTEGAELQLKEVTLGVATAPQOVLNSGVPSN 926
QY 1202 FTYSPDVKLDICVSLMGEQFLVSW 1227
DB 927 FTYSPDNKSLAIPVSLMGEFLQISW 952

RESULT 8

ABM84316
ID ABM84316 standard; protein; 788 AA.

AC ABM84316;

XX 18-NOV-2004 (first entry)

DE Human diagnostic and therapeutic pprotein SEQ ID NO:4565.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

OS WO2004023973-A2.

PN 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

PR 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Eider LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Geratin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Pollicky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Fatury S, Shi X, Suarez CJ;

XX WPI: 2004-329368/30.

DR N-PSDB; ACN42968.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, -in gene therapy or
PT in gene mapping.
XX Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIFO at www.wipo.int/pt/en/sequences/listing.htm

XX Sequence 788 AA;

Query Match 57.7%; Score 3763.5; DB 8; Length 788;
Best Local Similarity 79.0%; Pred. No. 1e-310;
Matches 724; Conservative 3; Mismatches 20; Indels 169; Gaps 2;
QY 313 LGRTRKEIGYVKKHLQDLSGRISARABAEATGAHPRPRAVPTQCDVPPNSRFFCAPDK 372
DB 42 LSGSSPVLSEETHPAHQOGASRPGPR-----DAQAHPRPRAVPTQCDVPPNSRFFCAPDK 96
QY 373 AITQOCEARGCCYIPAKQGLQAGQMPWCFPPSPYSYKLENLSSESSEMGYATLTRTT 432
DB 97 AITQOCEARGCCYIPAKQGLQAGQMPWCFPPSPYSYKLENLSSESSEMGYATLTRTT 156
QY 433 PTFPPKIDILRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEE 492
DB 157 PTFPPKIDILRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEE 216
QY 493 PRGVIVRQDGRVLLNTTVAFLPADOFLQSLPSQYITGLAEHLSPLMLSTSWTRI 552
DB 217 PRGVIVRQDGRVLLNTTVAFLPADOFLQSLPSQYITGLAEHLSPLMLSTSWTRI 276
QY 553 TLMNRDLAPTGANLYGSHPPFVLALEDGSAHGVLNLSNAMDVVLQSPALSWSTGGI 612
DB 277 TLMNRDLAPTGANLYGSHPPFVLALEDGSAHGVLNLSNAMDVVLQSPALSWSTGGI 336
QY 613 LDVYIFLGPPEKSVVQQYLDVVGYFMPYMGFLCRWGYSSTAITRQVVENNTRAHP 672
DB 337 LDVYIFLGPPEKSVVQQYLDVVGYFMPYMGFLCRWGYSSTAITRQVVENNTRAHP 396
QY 673 PLDVQWMDLDYMSRRDFTFNKDGPRDPPAMVQELHQGRRYMMIVDPAISSGGPAGSYR 732
DB 397 PLDVQWMDLDYMSRRDFTFNKDGPRDPPAMVQELHQGRRYMMIVDPAISSGGPAGSYR 456
QY 733 PYDEGLRRGVFITNETGOLIGKVPWGSTAPDFTNPTALAWMEDVAFHDOVDFDGLW 792
DB 457 PYDEGLRRGVFITNETGOLIGKVPWGSTAPDFTNPTALAWMEDVAFHDOVDFDGLW 516
QY 793 IDMBNPSFIRSGEDGCNNELNPPYVGVGGTLOAATICASSHQFLSTHYNLHLYG 852
DB 517 IDMBNPSFIRSGEDGCNNELNPPYVGVGGTLOAATICASSHQFLSTHYNLHLYG 576
QY 853 LTEATASHRALVKARCTPFFVISRSTFAGHGYAGHWTGDVSSWEOLASSVPEILOLNL 912
DB 577 LTEATASHRALVKARCTPFFVISRSTFAGHGYAGHWTGDVSSWEOLASSVPEILOLNL 636
QY 913 LCVPLVGADVCGFLGNTSEELCVMTQLGAFYPMNRNHNLSLSPQEPYSFSEPAQAMR 972
DB 637 LCVPLVGADVCGFLGNTSEELCVMTQLGAFYPMNRNHNLSLSPQEPYSFSEPAQAMR 696
QY 973 KALTURYALLPHLYTLFHOAHVAGETVARPLFLEPKDSSTWTDVHQLLMGALLITPVL 1032
DB 697 KALT----- 700
QY 1033 QAGKAEVTCYPLGTWYDLOTVPIEALGSLPPPPAPRAIPAHSEGQWVTLPAPLDTINV 1092
DB 701----- 700
QY 1093 HLRAGYIIPLOGPGLTTTESRQPPMALAVALTKGGEARGELFDWDDGESLEVLERGAYTOV 1152
DB 701-----LEVLERGAYTOV 712
QY 1153 IFLARNNTIIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVL 1212
DB 713 IFLARNNTIIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVPSNFTYSPDTKVL 772
QY 1213 ICVSLLMGEQFLVSWC 1228
DB 773 ICVSLLMGEQFLVSWC 788
RESULT 9
ADJ83134
ID ADJ83134 standard; protein; 707 AA.
XX XX

AC	ADJ83134;	XX	New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,
XX		PT	preventing or treating NOVX-associated polypeptide disorder, e.g.
DT	06-MAY-2004 (first entry)	PT	cardiomyopathy, atherosclerosis, diabetes, cancer, Parkinson's disease or
XX		PT	asthma.
DE	Glucosyl hydrolase family 31 protein - SEQ ID 125.	XX	
XX		PS	Disclosure; SEQ ID NO 125; 263pp; English.
KW	NOVX; cytostatic; antiarteriosclerotic; antidiabetic; antiasthmatic;	XX	The invention relates to a novel isolated NOVX polypeptide. The
KW	antiasthmatic; antiinflammatory; respiratory; antiarthritic;	CC	polypeptide demonstrates cytostatic, antiarteriosclerotic, antidiabetic,
KW	dermatological; antibacterial; cerebroprotective; vasotropic; cardiant;	CC	antiasthmatic, antiasthmatic, antiinflammatory, respiratory,
KW	haemostatic; hypotensive; hepatotropic; neuroprotective; anorectic;	CC	antiasthmatic, dermatological, antibacterial, cerebroprotective,
KW	nootropic; antilucer; muscular; immunosuppressive; gynaecological;	CC	vasotropic, cardiant, haemostatic, hypotensive, hepatotropic,
KW	antiparkinsonian; anticonvulsant; tranquiliser; analgesic; nephrotropic;	CC	neuroprotective, anorectic, nootropic, antilucer, muscular,
KW	antiinfertility; antilipemic; cardiomyopathy; atherosclerosis; diabetes;	CC	immunosuppressive, gynaecological, antiparkinsonian, anticonvulsant,
KW	cell signal processing; metabolic pathway; asthma; allergy; emphysema;	CC	ophthalmological, osteopathic, antiparkinsonian, anticonvulsant,
KW	obesity; graft-versus-host; arthritis; cancer; stroke; haemophilia;	CC	tranquilliser, analgesic, nephrotropic, antiinfertility and antilipemic
KW	activity; Alzheimer's; pain; chromosome mapping; tissue typing;	CC	activities. The NOVX polypeptide, nucleic acid or antibody of the
XX	glucosyl hydrolase family 31; enzyme.	CC	invention may be useful for treating or preventing a NOVX-associated
XX	Unidentified.	CC	disorder, such as cardiomyopathy, atherosclerosis, diabetes or a disorder
OS		CC	related to cell signal processing and metabolic pathway modulation.
XX	US2003170630-A1.	CC	Furthermore, the NOVX polypeptides may be useful for diagnosing, treating
PN		CC	or preventing diseases such as asthma, allergies, emphysema, autoimmune
XX	11-SEP-2003.	CC	disease, graft-versus-host disease, arthritis, cancer, stroke,
XX		CC	haemophilia, obesity, Alzheimer's disease and pain. The nucleic acids may
XX	21-DEC-2001; 2001US-00032189.	CC	be used as hybridisation probes, in chromosome mapping, tissue typing, of
XX		CC	a protein of the invention which is related to human NOVX protein.
PR	21-DEC-2000; 2000US-0257495P.	XX	Sequence 707 AA;
PR	22-DEC-2000; 2000US-0258171P.	Query Match	31.0%; Score 2024; DB 7; Length 707;
PR	20-FEB-2001; 2001US-0269940P.	Best Local Similarity	54.7%; Pred. No. 1.5e-162;
PR	08-MAR-2001; 2001US-0274192P.	Matches 399; Conservative	87; Mismatches 213; Indels 30; Gaps 9;
PR	22-MAR-2001; 2001US-0277826P.		
PR	29-MAR-2001; 2001US-0279840P.	QY	470 PLETRVHSRAEPLYSVEFSEPEGVIVHRQDGRVLTNTTVAPEADQFLQSLSLP 529
PR	11-APR-2001; 2001US-0282981P.	Db	1 PLDFPASSPASBDDYDLVNSGPFGEVIRKSTGDLVDFDTTFTGVLVDFDQFLQSLSLP 60
PR	13-APR-2001; 2001US-0283656P.	QY	530 SQVITGLAEHLSPLMLSTSWTRITLWNRDLATPG-ANLYGSHPPFYLALEDGGSARHGVFL 588
PR	31-JUL-2001; 2001US-0309247P.	Db	61 SEYIYGLGEHAKLFRDNTNETYTLWNRDVGYPYSGDNNLYGSHPPFYMSLEDGSAHGVFL 120
PR	10-AUG-2001; 2001US-0311754P.	QY	589 LNSNMDVVLQSPALSWRSTGGILDVYIFLQPEKSVVQQKLDVGVGPFMPYWGSLGFH 648
PR	17-AUG-2001; 2001US-0313331P.	Db	121 LNSNAMEVDIGEPALTYRVIGILDYFVFLGFTPEDVLQYTELIGRALPYWLSLGFH 180
XX	(ALSO/) ALSOBROOK J P.	QY	649 LCRWGYSSTAITROVVENMTRAHFFLDVQWDLDMDSRDEFTENKDGFRDPAMYQELH 708
PA	(TCH/) TCHERNEV V T.	Db	181 LCRWGYTNVSEVKTVDGMRKANIPLDVQWLMDIDYMDGYKDFDTPVRPFGPEDFVKKLH 240
PA	(LIUX/) LIU X.	QY	709 QGRRYMIIVDPAISSGPGAGSYRYPYDEGLRRGVFTTNETGQPLIGKVPWGSTAFDFTN 768
PA	(SPY/) SPYTEK K A.	Db	241 AKQKVVILDPASVD--SASYYPYERKEGVFKVKNPNSGSDYIGEVWPGYTAFFDFTN 298
PA	(PAT/) PATTURAJAN M.	QY	769 PTALAWEDMVAEFHQVFPDGLWIDMNEPSNFIKRGSEDCPNNELENPPYVGVGGTL 828
PA	(LEPL/) LEPLY D M.	Db	299 PEARKWAEIKDFHDSLPFDGIWIDMNEPSSF--SEPGNDSNLNLYPPYAPNDGDGPL 355
PA	(BUR/) BURGESS C B.	QY	829 QAATICASS-HQFLSTHYNLNLGLTEAISHRALVKAR-CTRPVVISRSTFAGHGRVA 886
PA	(SHIM/) SHINKETS R A.	Db	356 SSKTMCMDAVHYGVHVDVHNLGLSEAKAYEALKKVTGKRPVLSRSTFAGSGRVA 415
PA	(GROS/) GROSSE W M.	QY	887 GHWTGDVMSWEQLASSVPEILQFNLLGVPLVGADVCGFLGNTSELCVWRVQLGAFYFP 946
PA	(SZEK/) SZEKERES E S.	Db	416 GHWTGDNTASDDLLKYSIPGVLSFNLFGIPFGVADICGNGNTTELCVWRVQLGAFYFP 475
PA	(VERN/) VERNET C A M.	QY	947 MNHNSLLSLPQEPYFSFSEPAQAMRKALTLRYALLPHLYTLFQAHVAGETVARPLFLUE 1006
PA	(LILL/) LI L.	Db	476 SRNHNHLGTPQEPWLFDSVAEASRKALNLYTLPLLYTLFHEAHVSGLPVWRPLPFE 535
PA	(CASM/) CASMAN S J.	QY	1007 FPKDSTWTVDHQLLWGEALLITPVLOAKAEVTFYPLGTWYDLOTVEIALGSLPPPP 1066
PA	(BOLD/) BOLDOG F L.		
PA	(GORM/) GORMAN L.		
PA	(GANG/) GANGOLLI E A.		
PA	(FERN/) FERNANDES E R.		
PA	(RIEG/) RIEGER D K.		
PA	(EDIN/) EDINGER S R.		
PA	(GUNT/) GUNTHER E.		
PA	(MILL/) MILLET I.		
PA	(SCIO/) SCIORE P.		
PA	(ELLE/) ELLERMAN K.		
PA	(MACD/) MACDOUGALL J R.		
PA	(SMIT/) SMITHSON G.		
XX	Alsobrook JF, Tchernev VT, Liu X, Spytek KA, Zerhusen BD;		
PI	Patturajan M, Lepley DM, Burgess CE, Shimkets RA, Grosse WM;		
PI	Szekeres ES, Vernet CAM, Li L, Casman SJ, Boldog FL, Gorman L;		
PI	Gangolli EA, Fernandes ER, Rieger DK, Edinger SR, Gunther E;		
PI	Millet I, Sciore P, Ellerman K, Macdougall JR, Smithson G;		
XX	WPI; 2003-898249/82.		

Dd	536	FPDDAETDIDRQFLWGSALLVAPLEPGATSKAYLPGGRWYDLYTGAGEA-----	587	CC	the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm		
Qy	1067	AAPREPAIHSEGOWTLAPLDITNVHLRAGYIIPLOQGLTTTSSRQOPMALAVATKGG	1126	XX			
Dd	588	-----SRGNVTLSPDKIPVHVVRGGSIIPTQEPALTTTSSRDNPFHLLVALDDN	638	SQ	Sequence 638 AA;		
Qy	1127	GEARGELFWDDGESLEVLERGAYQVIFLARNNITVNELVRVT---SEGAGLQLOKQVTVL	1183		Query Match 30.8%; Score 2007.5; DB 8; Length 638; Best Local Similarity 49.7%; Pred. No. 3.2e-161; Matches 465; Conservative 19; Mismatches 94; Indels 357; Gaps 22;		
Dd	639	GTASGELYLDDGESIDT-QRGDYLVLVQFSANNNTLTGTGV-VTGYYKNSNTLTLEKITIL	696	Qy	313	LEGRTKELGYTVKKHLQDLSGRISARAEATGAHGRPRAPVPTQCDVPPNSRFDCAPDK	372
Qy	1184	GVATAPOQV 1192		Dd	42	LSGSSPVLTEETHPAHQCGASRGPR-----DAQAHGPRAPVPTQCDVPPNSRFDCAPDK	96
Dd	697	GVGNEPAAV 705		Qy	373	AITQEQCEARGCCYIPAKQGLQAGMQPWCFFPPSPYSYKLENLSSEMGYATATLRTT	432
				Dd	97	AITQEQCEARGCCYIPAKQGLQAGMQPWCFFPPSPYSYKLENLSSEMGYATATLRTT	156
				Qy	433	PTFFPKDILTLLRLDVMETENRLHFTIKDPANRRYEVPLETRVHSHRAPSPLYSVEFSEE	492
				Dd	157	PTFFPKDILTLLRLDVMETENRLHFTIKDPANRRYEVPLETRVHSHRAPSPLYSVEFSEE	216
				Qy	493	PGVIVRQDGRVLLNTTVAPLFEADQFLOLSTLPSQYITGLAEHLSPMLSTSWTRI	552
				Dd	217	PGVIVRQDGRVLLNTTVAPLFEADQFLOLSTLPSQYITGLAEHLSPMLSTSWTRI	276
				Qy	553	TLWNRDLAPTGCANLYGSHPPFYLALEDGSGAHGVFLNLSNAMDVVLQSPALSWRSTGGI	612
				Dd	277	TLWNRDLAPTGCANLYGSHPPFYLALEDGSGAHGVFLNLSNAMDVVLQSPALSWRSTGGI	336
				Qy	613	LDVYIFLGPPEKSVVQYLDVVVG-----YPMPPYVWGLGFHLCKRWGYSSTAITROV	663
				Dd	337	LDVYIFLGPPEKSVVQYLDVVVGACSLAAAPAPRLPP-----PSL	377
				Qy	664	VENMTFAHFPLDVQWNDLDYMSRRDFTFNKDGFRDFPAMVQELHOGGRRYMMIVDPAIS	723
				Dd	378	MKSALACRIPV-----HAAILGPGULP	398
				Qy	724	SSGPAG-----SYRPYDEGLRRGVFTNETGOLIGKVPWPGSTAFDPFTNPTALAWMEDV	779
				Dd	399	PV-PLGLLLHRYHP-----PGGGE-----	416
				Qy	780	AEFHQVQPFGLWIDMNEPSNFIRSGEDGCPNNELENPPYVGVVGGTILQAAATCASSHQ	839
				Dd	417	---HDQGLP-----PGRPVERP-----	431
				Qy	840	FLSTHYNLHNLVGLTEAISHRALVKARGTRPFFVISRSTFAGHGRYAGHWTGDVWSSWEQ	899
				Dd	432	-----GLHGLPEGL-----HVQQGW-----	446
				Qy	900	LASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWTLQLGAFYPFMRNHNLSLSLPQE	959
				Dd	447	-----LPGLP-----	451
				Qy	960	PVSFSEPAQAMRKALTURLYALLPHLYTLFHOAHVAGETVARPLFLEPKDSTWTVDHQ	1019
				Dd	452	-----GHGAGAAPGRPA-LHDDRGSC-----HQ	473
				Qy	1020	LLWGEALLITPVLOAGKAETGYF-----PLGTWYDLOTVPIEALGSLPPPPAAPREPA	1073
				Dd	474	QL-GPCREL-QALRRGSAB--GGFHQRDRPADMEGWARV-----HCLPRLHQ	518
				Qy	1074	IHSEQWVTLPAPLDTINVHLRAGYIIPLOQGLTTTSSRQOPMALAVATKGGARGEL	1133
				Dd	519	PHS-----PGL-----QPMALAVATKGGARGEL	543
				Qy	1134	FWDGESLEVLERGAYQVIFLARNTVIVNELVRVTSEGAGLQLOKQVTVLGVATAPQOVL	1193
				Dd	544	FWDGESLEVLERGAYQVIFLARNTVIVNELVRVTSEGAGLQLOKQVTVLGVATAPQOVL	603
				Qy	1194	SNGVPVSNFTYSPDKVLDICVSLLMGEQFLVSWC	1228
				Dd	604	SNGVPVSNFTYSPDKVLDICVSLLMGEQFLVSWC	638

Dd	536	FPDDAETDIDRQFLWGSALLVAPLEPGATSKAYLPGGRWYDLYTGAGEA-----	587	CC	the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm		
Qy	1067	AAPREPAIHSEGOWTLAPLDITNVHLRAGYIIPLOQGLTTTSSRQOPMALAVATKGG	1126	XX			
Dd	588	-----SRGNVTLSPDKIPVHVVRGGSIIPTQEPALTTTSSRDNPFHLLVALDDN	638	SQ	Sequence 638 AA;		
Qy	1127	GEARGELFWDDGESLEVLERGAYQVIFLARNNITVNELVRVT---SEGAGLQLOKQVTVL	1183		Query Match 30.8%; Score 2007.5; DB 8; Length 638; Best Local Similarity 49.7%; Pred. No. 3.2e-161; Matches 465; Conservative 19; Mismatches 94; Indels 357; Gaps 22;		
Dd	639	GTASGELYLDDGESIDT-QRGDYLVLVQFSANNNTLTGTGV-VTGYYKNSNTLTLEKITIL	696	Qy	313	LEGRTKELGYTVKKHLQDLSGRISARAEATGAHGRPRAPVPTQCDVPPNSRFDCAPDK	372
Qy	1184	GVATAPOQV 1192		Dd	42	LSGSSPVLTEETHPAHQCGASRGPR-----DAQAHGPRAPVPTQCDVPPNSRFDCAPDK	96
Dd	697	GVGNEPAAV 705		Qy	373	AITQEQCEARGCCYIPAKQGLQAGMQPWCFFPPSPYSYKLENLSSEMGYATATLRTT	432
				Dd	97	AITQEQCEARGCCYIPAKQGLQAGMQPWCFFPPSPYSYKLENLSSEMGYATATLRTT	156
				Qy	433	PTFFPKDILTLLRLDVMETENRLHFTIKDPANRRYEVPLETRVHSHRAPSPLYSVEFSEE	492
				Dd	157	PTFFPKDILTLLRLDVMETENRLHFTIKDPANRRYEVPLETRVHSHRAPSPLYSVEFSEE	216
				Qy	493	PGVIVRQDGRVLLNTTVAPLFEADQFLOLSTLPSQYITGLAEHLSPMLSTSWTRI	552
				Dd	217	PGVIVRQDGRVLLNTTVAPLFEADQFLOLSTLPSQYITGLAEHLSPMLSTSWTRI	276
				Qy	553	TLWNRDLAPTGCANLYGSHPPFYLALEDGSGAHGVFLNLSNAMDVVLQSPALSWRSTGGI	612
				Dd	277	TLWNRDLAPTGCANLYGSHPPFYLALEDGSGAHGVFLNLSNAMDVVLQSPALSWRSTGGI	336
				Qy	613	LDVYIFLGPPEKSVVQYLDVVVG-----YPMPPYVWGLGFHLCKRWGYSSTAITROV	663
				Dd	337	LDVYIFLGPPEKSVVQYLDVVVGACSLAAAPAPRLPP-----PSL	377
				Qy	664	VENMTFAHFPLDVQWNDLDYMSRRDFTFNKDGFRDFPAMVQELHOGGRRYMMIVDPAIS	723
				Dd	378	MKSALACRIPV-----HAAILGPGULP	398
				Qy	724	SSGPAG-----SYRPYDEGLRRGVFTNETGOLIGKVPWPGSTAFDPFTNPTALAWMEDV	779
				Dd	399	PV-PLGLLLHRYHP-----PGGGE-----	416
				Qy	780	AEFHQVQPFGLWIDMNEPSNFIRSGEDGCPNNELENPPYVGVVGGTILQAAATCASSHQ	839
				Dd	417	---HDQGLP-----PGRPVERP-----	431
				Qy	840	FLSTHYNLHNLVGLTEAISHRALVKARGTRPFFVISRSTFAGHGRYAGHWTGDVWSSWEQ	899
				Dd	432	-----GLHGLPEGL-----HVQQGW-----	446
				Qy	900	LASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWTLQLGAFYPFMRNHNLSLSLPQE	959
				Dd	447	-----LPGLP-----	451
				Qy	960	PVSFSEPAQAMRKALTURLYALLPHLYTLFHOAHVAGETVARPLFLEPKDSTWTVDHQ	1019
				Dd	452	-----GHGAGAAPGRPA-LHDDRGSC-----HQ	473
				Qy	1020	LLWGEALLITPVLOAGKAETGYF-----PLGTWYDLOTVPIEALGSLPPPPAAPREPA	1073
				Dd	474	QL-GPCREL-QALRRGSAB--GGFHQRDRPADMEGWARV-----HCLPRLHQ	518
				Qy	1074	IHSEQWVTLPAPLDTINVHLRAGYIIPLOQGLTTTSSRQOPMALAVATKGGARGEL	1133
				Dd	519	PHS-----PGL-----QPMALAVATKGGARGEL	543
				Qy	1134	FWDGESLEVLERGAYQVIFLARNTVIVNELVRVTSEGAGLQLOKQVTVLGVATAPQOVL	1193
				Dd	544	FWDGESLEVLERGAYQVIFLARNTVIVNELVRVTSEGAGLQLOKQVTVLGVATAPQOVL	603
				Qy	1194	SNGVPVSNFTYSPDKVLDICVSLLMGEQFLVSWC	1228
				Dd	604	SNGVPVSNFTYSPDKVLDICVSLLMGEQFLVSWC	638

Human diagnostic and therapeutic pprotein SEQ ID NO:4566.

gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

Homo sapiens.

WO2004023973-A2.

25-MAR-2004.

12-SEP-2003; 2003WO-US028227.

12-SEP-2002; 2002US-0410259P.

12-SEP-2002; 2002US-0410260P.

(INCY-) INCYTE CORP.

Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F; Harthshorne TA, Suchorolski MT, Altus CM, Pitts SU, Elder LV; Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP; Stevens KA, Blanchard JU, Panzer SR, Wang X, Au AP, Gerstin EH; Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LU; Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES; Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JU, Gietzen D; Fatury S, Shi X, Suarez CJ; WPI; 2004-329368/30. N-PSDB; ACN42969.

New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.

Claim 27; Page; 190pp; English.

The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2722 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, autoimmune/inflammatory disorder, developmental disorder, endocrine disorder, neurological disorders, gastrointestinal disorders, or infections caused by virus, bacteria, fungi or parasite. The dithp molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline gene therapy. The present sequence represents a dithp protein of the invention. Note: The sequence data for this patent is not represented in

ABM84317	ABM84317	standard; protein; 638 AA.
AC	ABM84317;	
DT	18-NOV-2004	(first entry)
DE	Human diagnostic and therapeutic pprotein SEQ ID NO:4566.	
XX	gene therapy; human diagnostic and therapeutic polynucleotide; dithp.	
XX	Homo sapiens.	
XX	WO2004023973-A2.	
XX	25-MAR-2004.	
XX	12-SEP-2003; 2003WO-US028227.	
XX	12-SEP-2002; 2002US-0410259P.	
PR	12-SEP-2002; 2002US-0410260P.	
XX	(INCY-) INCYTE CORP.	
XX	Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F, Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV, Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP, Stevens KA, Blanchard JL, Paneser SR, Wang X, Au AP, Gerstin EH, Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL, Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES, Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D, Patury S, Shi X, Suarez CJ;	
XX	WPI; 2004-329368/30.	
DR	N-PSDB; ACN42969.	
XX	New diagnostic and therapeutic polynucleotides and polypeptides, useful in diagnosing a condition, disease or disorder associated with human molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or in gene mapping.	
XX	Claim 27; Page; 190pp; English.	
XX	The invention relates to novel diagnostic and therapeutic polynucleotides selected from one of the 2722 sequences defined in the specification. A polynucleotide of the invention may have a use in gene therapy. The human diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, autoimmune/inflammatory disorder, developmental disorder, endocrine disorder, neurological disorders, gastrointestinal disorders, or infections caused by virus, bacteria, fungi or parasite. The dithp molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline gene therapy. The present sequence represents a dithp protein of the invention. Note: The sequence data for this patent is not represented in	

```

RESULT 11
AD019560
ID ADO19560 standard; protein; 1857 AA.
XX AC ADO19560;
XX DT 12-AUG-2004 (first entry)
XX DE Human PRO polypeptide #245.
XX KW Human; PRO; immune related disorder; systemic lupus erythematosus;
XX KW rheumatoid arthritis; osteoarthritis; juvenile chronic arthritis;
XX KW systemic sclerosis; Sjogren's syndrome; vasculitis; sarcoidosis;
XX KW autoimmune haemolytic anaemia; autoimmune thrombocytopenia; thyroiditis;
XX KW diabetes mellitus; renal disease; demyelinating disease;
XX KW central nervous system; peripheral nervous system;
XX KW demyelinating polyneuropathy; Guillain-Barre syndrome;
XX KW chronic inflammatory demyelinating polyneuropathy.
XX OS Homo sapiens.
XX PN WO2004043361-A2.
XX PD 27-MAY-2004.
XX PF 06-NOV-2003; 2003WO-US035268.
XX PR 08-NOV-2002; 2002US-0425235P.
XX PA (GETH ) GENENTECH INC.
XX PI Fong S, Dennis K, Clark H, Chiu H, Schoenfeld J, Williams PM;
XX PI Wood WI, Wu TD;
XX DR WPI; 2004-420067/39.
XX DR N-PSDB; AD019559.
XX PT Novel PRO polypeptide e.g., PRO69614, PRO71106, or PRO86388 useful for
XX PT treating an immune related disorder such as systemic lupus erythematosus,
XX PT rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis or
XX PT spondyloarthritis.
XX PS Claim 7; SEQ ID NO 490; 1731pp; English.
XX CC The invention relates to human PRO polypeptides and the polynucleotides
XX CC encoding them. The polypeptides and polynucleotides are useful for
XX CC treating and diagnosing immune related disorders in mammals. The immune
XX CC related disorders include systemic lupus erythematosus, rheumatoid
XX CC arthritis, osteoarthritis, juvenile chronic arthritis, systemic
XX CC sclerosis, Sjogren's syndrome, vasculitis, sarcoidosis, autoimmune
XX CC haemolytic anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes
XX CC mellitus, immune-mediated renal disease, demyelinating diseases of the
XX CC central or peripheral nervous system, demyelinating polyneuropathy,
XX CC Guillain-Barre syndrome and chronic inflammatory demyelinating
XX CC polyneuropathy. This sequence represents a human PRO polypeptide of the
XX CC invention.
XX SQ Sequence 1857 AA;

Query Match 29.1%; Score 1900.5; DB 8; Length 1857;
Best Local Similarity 43.7%; Pred. NO. 2.7e-151;
Matches 394; Conservative 133; Mismatches 328; Indels 47; Gaps 21;

QY 348 RGRPRVPTQCDVP---PNSRDCAPDKAITQCEARGCCYIPAKQGLQGMGPWCF 404
DB 78 PGTGTTPVSACPVVNERINCIPDPPTKATCDQRCGWNP-----QGA-VSPWCY 131
QY 405 FPPSPSYKLE-NLSSEMGYATLTRTPT--FPPKDLTLRLDVMVMTENRLHPTKD 461
DB 132 YSKNH-SYHVGMLVNTAGTARL-KNLPSPVFGSDVNDVLLTAYQTSNRHFPLTD 189
QY 462 PANRRYVEPLETRVHS--RAPSPLYSVFSEEPFGVIVHRQLDGRVLLTTTVAFLPFAD 519

```


PS Disclosure; Page 114-119; 119pp; English.

CC The invention relates to in vitro screening of metastatic colorectal cancer cells or primary and/or metastatic stomach or oesophageal cancer cells by testing cells in extra-intestinal tissues and/or body fluids for expression of SI (sucrase isomaltase), CDX1 or CDX2 (transcription factors). Expression of these markers indicates possible presence of the specified cancer cells. The method is used to diagnose (or monitor) metastatic colorectal cancer or primary and/or metastatic stomach or oesophageal cancer cells, also to confirm identification of such cells. These cancers can be treated by administration of an SI ligand and (optionally conjugated) cytostatic agent or radioimaged by administering a conjugate of the SI ligand and detectable agent. The present sequence represents human SI

XX

SQ Sequence 1827 AA;

Query Match 28.8%; Score 1877.5; DB 4; Length 1827;

Best Local Similarity 41.7%; Pred. No. 2.4e-149;

Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

QY 340 AEAEAGHGRPRRAVPTQCDVPNPSRFDCAIDKAITQEQCEARGCCVPAKQGLQGAQMG 399

DB 49 ATTRVTNPSDSKCPNVLNDPVNVRINCIPQEPTEGICAQGCCWRPNWDSL----- 102

QY 400 QPWCFFPPSPYSKLENLSSEMGYATLTR--TTPTFFPKDILTLRLDVMETENLHFT 458

DB 103 IPWCFFVDNH-GYNVDQMTTTSIGVEAKLRIPSLTFLGNDINSVLFTTQNTQPNRFRK 161

QY 459 IKDPAARRVEVPLETRVHS--RAPPLYSVESEPPGVIVHRQLDGRVLLNTVAPLF 516

DB 162 ITDPNRRYEVPHQYKFEKFTGPTVSTLDYDKVAQNPFSIQVIRKSGKTLFTSGLPV 221

QY 517 FADQFLQLSLPSQYITGLAEHL-SPLMLSTGTRITLWNRDLAP--TPGANLYGSHPFY 574

DB 222 YSDQYLIQISARLSDIYIGEQVHKFRPHDLKSWKTPFTRQLPDGNNNLYGHQTFP 281

QY 575 LALED--GGSAGHGVFLNNSAMVDVLPSPALSWRSTGGILDVYIFLGPPEKSVVQYQYLDV 633

DB 282 MCIEDTSGKSGFVFLMNSNAMEIFIQTPITVYRVGTGGILDFYLLGDTPEQVQVQYQQL 341

QY 634 VGYPPFPWYGLFHLRCWGYSSAITRQVVENWTRAHPLDYNWDLVMDSRDFTEN 693

DB 342 VGLPAMPAYNNVQLFQSRMNYKSLDVVKEVVRNREAGIPFDQVTDIDYMEDKDYTD 401

QY 694 KDGFRDPPAMVQSLHOGGRYMMIVDPAISSGPG--SYRPYDEGLRGVFTTNETGQ- 750

DB 402 QVAFNGLPQFVQDLHDHGQKXVILDPALISIGRANGTYATYERGNTQHVWLNESDGT 461

QY 751 PLIGKVPGSTAPDFTNPTALAWEDMVAEPHDQVPFDGLWIDMNEPSNFIKGSBDGCP 810

DB 462 PIIGVMPGLTVYPDFTNPNCIDWANECSIFHQEVQYDGLWIDMNEVSSFIQGSTKGCN 521

QY 811 NNELENPVYGVVGTQLQATTCASHQPLSTHYNLNLYLGTETAEASHRALVKA-RGT 869

DB 522 VNKLANTPPTPDLDKMLYKTKTICMDAVQNMGRQYDVHSLYGSMAIATEAQVQKFPNK 581

QY 870 RPPVKSRTFAGHGVAGHTGVDWSSQELASSVPEILQNLGVLPLVCAQVCGFLGNT 929

DB 582 RSFILTRSTPAGGRHAHWLGONTASQEWESITGMLEFSLFGPLVGLADICGVVET 641

QY 930 SEELCVRTQLGAFYFMRNHNLSLIPQEPYFSBPA-----QOAMKRLTLRYAL 981

DB 642 TEELCRWMLQGAFFYFMRNHNLS-----DGEHQDPAPFGQNSLLVKSRQVLTIRYL 695

QY 982 LPHLYTLFHOAHVAGETVAPPLFEPKDSSTVVDHLLWGALLITPVLQKAEVTVG 1041

DB 696 LPFLYTLFKARHVGETVAPPLFHEFYEDTNSIEDTEFLWGPALLITPVLKQADTVSA 755

QY 1042 YPFLGTYDLOTPVIEALGSLPPPPAPREPAIHSQWTLTAPLDTTNVHLRAGVIIP 1101

DB 756 YIPDAIWDYDES-----GAKRP-----WRKQRVDMYLPADKIGLHLRGYIIP 798

QY 1102 LQPGGLTTTTSRQOPMALAVALTKGARGELFWDDGESLEVLERCAYTOVIFLARNTTI 1161

DB 799 IQEPDVTTATSRKNPLGLLIVAGENNATKGDFFWDDGETKOTIQNGYILYTFVSNNNTL 858

QY 1162 VNELVRVTS---EGAGLQKQKTVTLGV--ATAPOQVLSNGVPV---SNFTYSPDTKVLDI 1213

DB 859 --DIVCTHSSYQEGITLAFQTVKILGLTSDVTEVRVAENNPMPNAHSNFTYDASQVLLI 916

QY 1214 C-VSLLMGEQFLVSW 1227

DB 917 ADLKLNLGRNFSVQW 931

RESULT 14

ABGI4994

ID ABGI4994 standard; protein; 1829 AA.

XX

AC ABGI4994;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #14985.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR N-PSDB; AAS79181.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX

PS Claim 20; SEQ ID NO 45353; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

```
SQ Sequence 1829 AA;
Query Match      28.8%; Score 1877.5; DB 4; Length 1829;
Best Local Similarity 41.7%; Pred. No. 2.4e-149;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

QY 340 ABAETGAHGRPRAVPTQCDVPNSRFDPCAPDKAITQEOCEARGCCYIPAKOGLQAGMG 399
DB 51 ATRVTTNFSDSKCNVLNDPVNVRINCIPQEPFTEGICACQGCCWCPWNSL----- 104
QY 400 QWCFPPPPSPYKLENLSSEMGYATATR-TTTFPPKIDILTLRLDMMETENRLHFT 458
DB 105 IPWCFPPVDNH-GYNVQDDMTTISIGVEAKLNRPSPTLFGNDINSVLTFTQNTQNRPREK 163
QY 459 IKDPANRRYEVPLETFRVHS--RAPSPLYSVFSEBPPGVIVHRLDGRVLLNTTVPPLF 516
DB 164 IYDNNRRYEVPHQYKFTGTPTVSTLDVKAQNPFSIQVIRKSGKTLFDTSIGPLV 223
QY 517 FADQFLQLSTSLPSQYITGLAEHL-SPLMLSTSWTRITLWNRDLAP-TPGANLYGSHPPY 574
DB 224 YSDQYLQISARLPDSYIYIGEQVHKRFRHDLWSKWTPIFTRDQLPGDNNNLYGHQTF 283
QY 575 LALED-GGSAHGVLNLSAMVVLQSPALSWRSTGGILDVYIFLGPEPKSVVQOYLDV 633
DB 284 MCIEDTSKSGKSGVFLMNSNAMEIFIQPTIVTRYVTGGILDFYILLGDTPEQVVOYQOL 343
QY 634 VGVPEMPYVWGLGFHLCRGYGSTAITROVENMTTRAHPFLDVQWMDLDYMDSRDFTEN 693
DB 344 VGLPAMPATWNLGFLSRNYSKLDVVKEVRNRREAGIPFTQVTDIDYMDKDFDFTVD 403
QY 694 KQGFDFPAMVQELHGGRRYMMVDPALSSGPGAG--SYRPPYDEGLRGVFTITNMG- 750
DB 404 QVAFNGLPQVQDLHDHGQKVIILDPALISIGRRANGTTVATYRGNTQHVWINESDGT 463
QY 751 PLICKWPGSTAPPDNTNTALAWEDMVAEHDQVFPDGLMIDMNEPSNFRSGDGC 810
DB 464 PIIGYVMPGLTVYDPDTNFCINDWANECSIFHQEVQYDGLMIDMNEVSFISQSTKCN 523
QY 811 NNELENPVYVGVGTQAAATTCASHQFLSTHYNLNLYGLTEATASHRALVKA-RGT 869
DB 524 VNKLYNPTPTDILDKMLYSKIICMDAVQWKGQYDVHSLYGYSMATIEAQVKVFNK 583
QY 870 RPFVSRSTFAGHRYAGHTGVDWSSWQLASSVPEILQFNLGLVPLVGVADVCGFLNT 929
DB 584 RGFILTRSTFAGSGRAHAWLGDNTASWQMEWSITGMLEFSLFGIPLVGVADICGFV 643
QY 930 SEELCVRTOLGAFYFPMRHNLSLLSPQEPYSFSEPA-----QQAMRKALITRYAL 981
DB 644 TEELCRNMQLGAFYFPSRNHNS-----DGVEHODPAPFGQNSLLVKSRSQVLTIRYL 697
QY 982 LPHLYTLFHOAHVAGTVARPLFEPKDSSTWVDHQLWGEALITPVLQAGKAEVTG 1041
DB 698 LPPLYTLFKAHVGETVARPVLUHEFYEDTNSWIEDTEFLWGPALLITPVLKQADTVSA 757
QY 1042 YFPLGTWYDLQTVPIBALGSLPPPPAAPREPAIHSEGOVWTPLPAPLDTINVHLRAGYIIP 1101
DB 758 YIPDAIWDYDES-----GAKRP-----WRKQVDMYLPADKIGLHLRGGYIIP 800
QY 1102 LQPGPLTTTSTRQOPMALAVALTKGEARGELFWDDGESLEVLERGAYQVIFLARNTI 1161
DB 801 IQEPDVTITASRKNPLGLIYALGENTNATKGFDFWDDGETKDTIQNGNYITLYTFSVSNNTL 860
QY 1162 VNELVRVTS---EGAGLQKQVTLGV--ATAPOQVLSNGVPV---SNFTYSPDVKLDI 1213
DB 861 --DIVCTHSYQEGTTLAFQTVKILGLTDSVTEVRVAENNPQMAHNSFTYDASNOVLLI 918
QY 1214 C-VSLMNGSQFLVSW 1227
DB 919 ADLKLNLGRNFSVQW 933
```

RESULT 15
AAB48844

```
ID XX AAB48844 standard; protein; 357 AA.
AC XX AAB48844;
DT XX 13-MAR-2001 (first entry)
XX XX Human RAP (receptor associated protein).
DE XX RAP; receptor associated protein; human; factor VIII clearance;
KW KW LRP-mediated plasma clearance; receptor-dependent clearance;
KW KW receptor-independent clearance; ligand internalisation;
XX XX low density lipoprotein related protein; haemophilia; half-life.
OS XX Homo sapiens.
XX XX WO200071714-A2.
XX XX 30-NOV-2000.
XX XX 24-MAY-2000; 2000WO-US014111.
XX XX 24-MAY-1999; 99US-0135847P.
XX XX (AMNA-) AMERICAN NAT RED CROSS.
PI XX Saenko EL, Strickland DK;
XX XX WPI: 2001-025163/03.
XX XX N-PSDB; AAC48844.
XX XX Factor VIII mutants having increased half-life useful for treating
PT PT hemophilia, comprise one or more amino acid substitutions in the A2
PT PT and/or C2 domain of factor VIII.
XX XX Claim 59; Fig 4; 121pp; English.
PS XX The invention relates to human factor VIII mutants comprising an amino
XX XX acid substitution at one or more positions in the A2 domain and/or an
CC CC amino acid substitution at one or more positions in the C2 domain. The
CC CC invention also encompasses a factor VIII mutant which lacks a B domain
CC CC (AA848842). The factor VIII mutants have an increased half-life in the
CC CC bloodstream. The A2 domain mutants exhibit reduced LRP-dependent
CC CC (receptor-dependent) clearance of factor VIII, while C2 domain mutants
CC CC have reduced receptor-independent clearance. The invention also relates
CC CC to a method of using RAP (receptor associated protein), a protein which
CC CC inhibits LRP (low density lipoprotein related protein)-mediated ligand
CC CC internalisation, to increase the half-life of factor VIII. The mutant
CC CC factor VIII proteins, and nucleotides encoding them, are useful for
CC CC treating haemophilia. RAP, LRP-binding RAP mutants or fragments, and
CC CC nucleic acids encoding them may also be used in the treatment of
CC CC haemophilia, in combination with a mutant factor VIII protein or DNA of
CC CC the invention. The invention provides means of increasing the half-life
CC CC of factor VIII by reducing its clearance from plasma. The present
XX XX sequence represents human RAP
XX XX Sequence 357 AA;

Query Match      25.5%; Score 1660.5; DB 4; Length 357;
Best Local Similarity 95.1%; Pred. No. 4.4e-132;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

QY 1 MRGSPGALWLLAL-----RTVLGYSREKNQPKSPKRESGEPRMEKLNQWEKAQRL 55
DB 10 LAGLPALLLLLLFLGPPAASHGKGYREKNQPKSPKRESGEPRMEKLNQWEKAQRL 69
QY 56 HLPVRLAEHLADIKIQRDELAWKLLKLDGDEGEKEARLIRNLINVLAKYGLDGK 115
DB 70 HLPVRLAEHLADIKIQRDELAWKLLKLDGDEGEKEARLIRNLINVLAKYGLDGK 129
QY 116 ARQVTSNLSGTOEDGLDDPRLEKLVHAKTSKFSGBELDKLWREFLHHKEKVHYNVL 175
DB 130 ARQVTSNLSGTOEDGLDDPRLEKLVHAKTSKFSGBELDKLWREFLHHKEKVHYNVL 189
```

Qy	176	LETLSRTEIHNVI	SPSDLS	DIKGSV	LHSRHT	ELKEKLS	INOG	LDRLRR	VSHQ	GYS	TE	235
Db	190	LETLSRTEIHNVI	SPSDLS	DIKGSV	LHSRHT	ELKEKLS	INOG	LDRLRR	VSHQ	GYS	TE	249
Qy	236	AEFEPRVIDLW	DLAQ	ANLTD	KELEAF	REELKH	FEAKIE	KHNHYQ	KLEIA	HEKLR	HAE	295
Db	250	AEFEPRVIDLW	DLAQ	ANLTD	KELEAF	REELKH	FEAKIE	KHNHYQ	KLEIA	HEKLR	HAE	309
Qy	296	SVGDGERVSR	SRREK	HALL	EGRT	KELGY	TVKKH	LQDL	SGRIS	RAR		339
Db	310	SVGDGERVSR	SRREK	HALL	EGRT	KELGY	TVKKH	LQDL	SGRIS	RAR		353

Search completed: April 5, 2005, 01:02:24
Job time : 140 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2005, 22:59:17 ; Search time 55 Seconds
(without alignments)
1666.710 Million cell updates/sec

Title: US-10-600-862A-7
Perfect score: 6522
Sequence: 1 MRGPGALWLLALRTVLGS.....KVLDCVSLLMGEQFLVSWC 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata1/iaa/6A COMB.pep.*
4: /cgn2_6/prodata1/iaa/6B COMB.pep.*
5: /cgn2_6/prodata1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/prodata1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4744.5	72.7	952	4	US-09-949-016-5894
2	4744.5	72.7	960	4	US-09-949-016-9527
3	1877.5	28.8	1827	4	US-09-443-780C-14
4	1877.5	28.8	1827	4	US-09-079-723-179
5	1650	25.3	323	1	US-07-992-827D-1
6	1650	25.3	323	1	US-08-216-593-1
7	1650	25.3	323	5	PCT-US93-12380-1
8	1403.5	21.5	877	1	US-08-430-925A-4
9	1322	20.3	960	4	US-09-248-796A-16360
10	1314.5	20.2	682	4	US-09-744-926-2
11	1083.5	16.6	1003	4	US-09-949-016-10627
12	1061	16.3	209	1	US-07-992-827D-1
13	1061	16.3	209	5	PCT-US93-12380-2
14	1054	16.2	209	1	US-08-216-593-2
15	906	13.9	914	3	US-09-437-054A-8
16	891.5	13.7	919	3	US-09-437-054A-17
17	852.5	13.1	729	4	US-09-248-796A-17121
18	786.5	12.1	956	4	US-09-270-767-44032
19	751	11.5	205	1	US-07-992-827D-5
20	746	11.4	203	1	US-08-216-593-5
21	746	11.4	203	5	PCT-US93-12380-5
22	729.5	11.2	938	4	US-08-897-843A-1
23	664.5	10.2	693	4	US-09-376-343-2
24	515	7.9	275	4	US-09-248-796A-14129
25	509.5	7.8	788	4	US-09-489-039A-12831
26	485	7.4	449	4	US-09-248-796A-14130
27	422.5	6.5	1066	2	US-08-633-770A-1

ALIGNMENTS

RESULT 1

US-09-949-016-5894
; Sequence 5894, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5894
; LENGTH: 952
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-5894

Query Match	72.7%	Score	4744.5	DB	4	Length	952
Best Local Similarity	97.2%	Pred. No.	0				
Matches	890	Conservative	3	Mismatches	18	Indels	5
Gaps	1						
QY	313	LSGRKEIGYTVKKHLQDL	SGRISARA	AEATGAHPR	AVPTQCDVPPNSR	FDCAPDK	372
DB	42	LSGSPVL	EEHPAQ	CGASR	GPGR-----	DAQAHPGR	AVPTQCDVPPNSR
QY	373	AITQCCARGCCYIPAKQ	LGQAQMGQPC	FPFPPSPYSYK	LENLSSSS	EMGYTATLTRIT	432
DB	97	AITQCCARGCCYIPAKQ	LGQAQMGQPC	FPFPPSPYSYK	LENLSSSS	EMGYTATLTRIT	156
QY	433	PTFFPKDILTLRLD	VMMETENR	LHFTIKDPANRR	VEVPLETPRVHSR	APSPLYSVF	FSSE 492
DB	157	PTFFPKDILTLRLD	VMMETENR	LHFTIKDPANRR	VEVPLETPRVHSR	APSPLYSVF	FSSE 216
QY	493	PRGVIVHQLDGRVLTNT	TVAPL	FADQFLQ	LSIPSOYITGL	AHLSPML	STWTRI 552
DB	217	PRGVIVHQLDGRVLTNT	TVAPL	FADQFLQ	LSIPSOYITGL	AHLSPML	STWTRI 276
QY	553	TLWNRLDAPTPCANLY	GSHPPFLA	LEDGSGAHGV	FLLNSN	MDVVLQ	SPALSWRSTGCI 612
DB	277	TLWNRLDAPTPCANLY	GSHPPFLA	LEDGSGAHGV	FLLNSN	MDVVLQ	SPALSWRSTGCI 336
QY	613	LDVYIFLGP	PKSVVQV	LDVVGV	PFMPYWG	LFLCRW	GYSTAITRQVVENMTRAHF 672

Db 337 LDVYIFLGPEPKSVVQQYLDVVGVPEMPYPWGLGFHLCRWGYSSTAITRQVVENMTRAHF 396
Qy 673 PLDVQWNLDDYMSRRDFTFNKDGFRDFPAMVQELHOGGRRYMMIVDPAISSGPGAGSYR 732
Db 397 PLDVQWNLDDYMSRRDFTFNKDGFRDFPAMVQELHOGGRRYMMIVDPAISSGPGAGSYR 456
Qy 733 PYDEGLRRGVFIINETGQPLIGKVPFGSTAFDPFTNPTALAWMEDVMAEFHDOVPPDGLW 792
Db 457 PYDEGLRRGVFIINETGQPLIGKVPFGSTAFDPFTNPTALAWMEDVMAEFHDOVPPDGLW 516
Qy 793 IDMNPSNFIRSGEDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 852
Db 517 IDMNPSNFIRSGEDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 576
Qy 853 LITEAISHRALVKARGTRFPVISRSTFAGHGRYAGHTGVDWSSWQLASSVPEILQFNL 912
Db 577 LITEAISHRALVKARGTRFPVISRSTFAGHGRYAGHTGVDWSSWQLASSVPEILQFNL 636
Qy 913 LGVPLVGADVCGFLGNTSEELCVRWTFQAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 972
Db 637 LGVPLVGADVCGFLGNTSEELCVRWTFQAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 696
Qy 973 KALTLYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 1032
Db 697 KALTLYALLPHLYTLFHOAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 756
Qy 1033 QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOVWTLPAPLDTINV 1092
Db 757 QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOVWTLPAPLDTINV 816
Qy 1093 HLRAGYIIPLOGPGLTTSRQOPMALAVALTGKGARGELFWDGSELEVLERGAYTOV 1152
Db 817 HLRAGYIIPLOGPGLTTSRQOPMALAVALTGKGARGELFWDGSELEVLERGAYTOV 876
Qy 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLSNGVPSNFTYSPDTPKVL 1212
Db 877 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLSNGVPSNFTYSPDTPKVL 936
Qy 1213 ICVSLLMGEQFLVSWC 1228
Db 937 ICVSLLMGEQFLVSWC 952

RESULT 2
US-09-949-016-9527
; Sequence 9527, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9527
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9527

Query Match 72.7%; Score 4744.5; DB 4; Length 960;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

Qy 313 LEGRTEKELGYTKKHLQDLISGRISRARAEAEAGHPRPRAVPTQCDVPPNSRFDCAPK 372
Db 50 LSGSSPVLEETHPAHQGASRGPGR-----DAQAHGPRPRAVPTQCDVPPNSRFDCAPK 104
Qy 373 AITQOCCBARGCCYIPAKOGLQGAOMGOWPCFPFPSPYSYKLENLSSSMGYSATLITRTT 432
Db 105 AITQOCCBARGCCYIPAKOGLQGAOMGOWPCFPFPSPYSYKLENLSSSMGYSATLITRTT 164
Qy 433 PTFEPKDIILTLRLDVMETENELHFTIKDPANRRYEVEPLETPRVHSRAPSPLYSEFSE 492
Db 165 PTFEPKDIILTLRLDVMETENELHFTIKDPANRRYEVEPLETPRVHSRAPSPLYSEFSE 224
Qy 493 PFGVTVHRQLDGRVLLNTTVAFLPFADQFLQSLSTSLPSQYITGLAEHLSPLMLSTSWTRI 552
Db 225 PFGVTVHRQLDGRVLLNTTVAFLPFADQFLQSLSTSLPSQYITGLAEHLSPLMLSTSWTRI 284
Qy 553 TLWNRDLAPTGANLYGSHPPFYLALEDGGSAGHVFLNLSNMDVVLQSPALSWRSTGGI 612
Db 285 TLWNRDLAPTGANLYGSHPPFYLALEDGGSAGHVFLNLSNMDVVLQSPALSWRSTGGI 344
Qy 613 LDVYIFLGPEPKSVVQQYLDVVGVPEMPYPWGLGFHLCRWGYSSTAITRQVVENMTRAHF 672
Db 345 LDVYIFLGPEPKSVVQQYLDVVGVPEMPYPWGLGFHLCRWGYSSTAITRQVVENMTRAHF 404
Qy 673 PLDVQWNLDDYMSRRDFTFNKDGFRDFPAMVQELHOGGRRYMMIVDPAISSGPGAGSYR 732
Db 405 PLDVQWNLDDYMSRRDFTFNKDGFRDFPAMVQELHOGGRRYMMIVDPAISSGPGAGSYR 464
Qy 733 PYDEGLRRGVFIINETGQPLIGKVPFGSTAFDPFTNPTALAWMEDVMAEFHDOVPPDGLW 792
Db 465 PYDEGLRRGVFIINETGQPLIGKVPFGSTAFDPFTNPTALAWMEDVMAEFHDOVPPDGLW 524
Qy 793 IDMNPSNFIRSGEDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 852
Db 525 IDMNPSNFIRSGEDGCPNNELENPYPVGVGGTLQAATICASSHQFLSTHYNLHNLXG 584
Qy 853 LITEAISHRALVKARGTRFPVISRSTFAGHGRYAGHTGVDWSSWQLASSVPEILQFNL 912
Db 585 LITEAISHRALVKARGTRFPVISRSTFAGHGRYAGHTGVDWSSWQLASSVPEILQFNL 644
Qy 913 LGVPLVGADVCGFLGNTSEELCVRWTFQAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 1032
Db 645 LGVPLVGADVCGFLGNTSEELCVRWTFQAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 764
Qy 1033 QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOVWTLPAPLDTINV 1092
Db 765 QAGKAEVTGYFPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGOVWTLPAPLDTINV 824
Qy 1093 HLRAGYIIPLOGPGLTTSRQOPMALAVALTGKGARGELFWDGSELEVLERGAYTOV 1152
Db 825 HLRAGYIIPLOGPGLTTSRQOPMALAVALTGKGARGELFWDGSELEVLERGAYTOV 884
Qy 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLSNGVPSNFTYSPDTPKVL 1212
Db 885 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLSNGVPSNFTYSPDTPKVL 944
Qy 1213 ICVSLLMGEQFLVSWC 1228
Db 945 ICVSLLMGEQFLVSWC 960

RESULT 3
US-09-443-780C-14
; Sequence 14, Application US/09443780C
; Patent No. 6699973
; GENERAL INFORMATION:
; APPLICANT: O'Mahony, Daniel J
; APPLICANT: Seveso, Michela
; TITLE OF INVENTION: Antibodies To Peptides That Target G1T Receptors And Related Meth

```

Db      799 IQBPDVTTTASRKNPGLGLVALGENTAKGDFPDDGGETKDTTONGNYILYTSVSNNTL 858
Qy      1162 VNELVRVTS---EGAGLQKQKVTVLGV--ATAFQQVLSNGVPV---SNFTYSPTDKVLDI 1213
Db      859 -DIVCTHSSYQSGTTTAFQTKILGLTSDSVTEVRVAENNQPMNAHSNFTYDASNQVLLI 916
Qy      1214 C-VSLMGEOFLVSW 1227
Db      917 ADLKMLGRNFSVQW 931

RESULT 4
US-09-079-723-179
; Sequence 179, Application US/09079723
; Patent No. 6703362
; GENERAL INFORMATION:
; APPLICANT: Alvarez, Vernon L.
; APPLICANT: O'Mahony, Daniel J.
; APPLICANT: Lambkin, Imelda J.
; APPLICANT: Singleton, Judith
; APPLICANT: Patterson, Catherine A.
; APPLICANT: Cagney, Gerard M.
; APPLICANT: Belinka, Benjamin A.
; APPLICANT: Carter, John M.
; TITLE OF INVENTION: RANDOM PEPTIDES THAT BIND TO GASTRO-
; TITLE OF INVENTION: INTESTINAL TRACT (GIT) TRANSPORT RECEPTORS AND RELATE
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,723
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mierock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 179:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1827 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-09-079-723-179

Query Match 28.8%; Score 1877.5; DB 4; Length 1827;
Best Local Similarity 41.7%; Pred. No. 3.1e-153;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

Qy      340 ABAETGAHGRPRRAVPTQCDVPNSRFDCAPDKAITOECEARGCCYIPAKOGLOGAQM 399
Db      49 ATTRVTNPSDSKCPNVLNDVNVNRINCPISQPFTEGICQAQGCCRPWNDSL----- 102
Qy      400 QPWCRRPPSPYSKYLENLSSSSMGVYATLTR--TTTFPPKDLTLRLDVMMETENRLHFT 458
Db      103 IPWCFFVDNH-GYNQDMTTTISIGVEAKLNRIPSPTLFGNDINSVLFTTQNTQPNRFRFK 161
Qy      459 IKDPANRRYVEPLETPRVHS--RAPSPLYSVBFSEPFQVIVHRRQJGDRVLNLTTVAPLF 516

```

Db 162 ITDNNRRVEVPQYKFGPTVSTLDYDVKVAQNPFSIQVIRKSNKTLFDTSIGPLV 221
QY FADQFLQLSTSLPSQVITGLAHL-SPLMLSTSWTRITLWNRDLAP-TPCANLYGSHPPY 574
Db 222 YSDQYLQISARLPSDIYIGIGEOVHRFRHDLWSKTPITRDQLPDGNNNLYGHQTF 281
QY LALED-GGSAGHGVFLNSNAMVVLQSPALSWSRSTGGILDVYIFIGPEPKSVVQOYLDV 633
Db 282 MCIEDTSGKSGFVFLMNSNAMEIFIQTPITVTVRVVTGGILDFYILLGDTPEQVQYQQL 341
QY 634 VGYPPMPYGLGFHLCRWCYSTAITQVVENMTAHPFLDVQWMDLDYMSRRDFTFN 593
Db 342 VGLPAMPAYWNLGFQLSRMNYKSLDVVKEVRRNREAGIPFDQVTDIDYMEDKQFTYD 401
QY 694 KCGFRDPFAMVQBLHCGRRYMMVDPALSSSPAG--SVRPYDEGLRGVFTTNETGQ- 750
Db 402 QVAFNGLPQFQVQDLHGHQYVILDPALISGRANGTYAYYERNGTQHWINESDGT 461
QY 751 PLIGKVPGSTAPDFPTNPTALAWMDVMAEHDQVFPDGLMIDMNEPSNFIRSGDGP 810
Db 462 PIIGVWPGLTVYPTFTNPNCDWANECSIFHQEVQYDGLMDMNEVSSFIQGSTKGCN 521
QY 811 NHELENPPYVGVVGTGLOAATICASHQFLSTHYLNHLYGLTEAIAASHRALVKA-RGT 869
Db 522 VNKLANYPFTPDILDKLMTYSKTCMDAVQWQKQYDVHSLYGYSMIAIATEQAVQKVPFNK 581
QY 870 RPFVISRSTFAGHGRVAGHWGTGWSWQOLASSVPEILQFNLLGVPLVADVCGLGNT 929
Db 582 RSPILTRSTFAGSGRAAHWGLDNTASWQMEWSITGMEFSLFGIPLVADICGFVAET 641
QY 930 SEELCVRTQLGAFYFPMRNHNSLLSLPOEPYSFSEPA-----QQAMRKALTIRYAL 981
Db 642 TEELCRNMQLGAFYFPMRNHNS-----DGEHQDPAFFGQNSLLVKSQVLTIRYTL 695
QY 982 LPHLYTLFQAHVAGTAVRPLFEPKDSSTVVDHQLLWGEALLITVPLQAGKAEVGT 1041
Db 696 LPFLYTLFYKAHVFGTAVRPLVHEFYDTSNIEDTEFLWGPALLITVPLKQGDVTSVA 755
QY 1042 YFPLGTYWDLQTVPIEALGSLPPPPAPAPREPAIHSQGWVTLPAPLDTTNVHLRAGYIIP 1101
Db 756 YIPDALWYDYES-----GAKRP-----WRKQVDMVLPADKIGLHLRGYIIP 798
QY 1102 LQSPGLTTESRQOPMALAVATKGEARGELFWDGDESLEVLERGAYTQVIFLARNTI 1161
Db 799 IQEPDVTITASRKNPLGLIVALGENTAKGDFWDDGETKDTIQNGNYLYLTFVSNNLT 858
QY 1162 VNELVRVTS---EGAGLQLOKVTVLGV--ATAQOVLNGVPV---SNFTYSPDTKVLDI 1213
Db 859 --DIVCTHSSYQEGTTLAFQTVKILGTDSTVEVRVAENNQPMNAHSNFTYDASNQVLLI 916
QY 1214 C-VSLLMGEOFLVSW 1227
Db 917 ADLKNLGRNFSVQW 931

RESULT 5

US-07-992-827D-1
; Sequence 1, Application US/07992827D
; Patent No. 5474766
; GENERAL INFORMATION:
; APPLICANT: Schwartz, Alan L., Bu, Guojun
; TITLE OF INVENTION: Methods And Compositions For
; TITLE OF INVENTION: Inhibition Of Hepatic Clearance Of Tissue-Type
; TITLE OF INVENTION: plasmidogen Activator
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bryan Cave
; STREET: 245 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10167-0034

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, high
MEDIUM TYPE: Density 1.44 MB storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: DOS Version 6.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/992.827D
FILING DATE: 18-DEC-1992
CLASSIFICATION: 514
PRIOR APPLICATION DATA: NONE
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 00108/059047
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 692-1800
TELEFAX: (212) 692-1900
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-992-827D-1

Query Match 25.3%; Score 1650; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 8.7e-135;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 YSREKQKPKPKRSGEFMEKLNQWKAQRLHLPVRLAELHADLKIQERDELAWK 80
Db 1 YSREKQKPKPKRSGEFMEKLNQWKAQRLHLPVRLAELHADLKIQERDELAWK 60
QY 81 KLKLDGLDDEKEARLIRNLNVLAKYGLDGKDKARQVTSNLSGTQDGLDDPRLEKL 140
Db 61 KLKLDGLDDEKEARLIRNLNVLAKYGLDGKDKARQVTSNLSGTQDGLDDPRLEKL 120
QY 141 WHKAKTSKFSGEELDKLWREFLHHKKEVHEYNVLETLSTRTTEEHENVISPSDLSIDK 200
Db 121 WHKAKTSKFSGEELDKLWREFLHHKKEVHEYNVLETLSTRTTEEHENVISPSDLSIDK 180
QY 201 SVLHSRHTLEKELASINQGLDRLRRVSHQGYSTAEFEPRVIDLWLAQSANLTKEL 260
Db 181 SVLHSRHTLEKELASINQGLDRLRRVSHQGYSTAEFEPRVIDLWLAQSANLTKEL 240
QY 261 EAFRELKHFEAKTEKHNHYQKLEIAHEKLRHAEVSGDGRVSRREKHALLEGTTEL 320
Db 241 EAFRELKHFEAKTEKHNHYQKLEIAHEKLRHAEVSGDGRVSRREKHALLEGTTEL 300
QY 321 GYTVKKHLQDLSGRISRAR 339
Db 301 GYTVKKHLQDLSGRISRAR 319

RESULT 6

US-08-216-593-1
; Sequence 1, Application US/08216593
; Patent No. 5650391
; GENERAL INFORMATION:
; APPLICANT: Broze, George J., Jr.
; APPLICANT: Schwartz, Alan L.
; APPLICANT: Warshawsky, Ilka
; TITLE OF INVENTION: Methods and Compositions for Inhibition
; TITLE OF INVENTION: of Hepatic Clearance of Tissue Factor Pathway Inhibitor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; ADDRESSEE: Corporate Patent Dept.
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA

QY 826 -----GTLQATTCASS----- 837
DB 509 SDYPLGDFKSNASEWKSISEAAATKTITTTSSASTSIDGKNTLAPKGNINYPYAIN 568
QY 838 -----HOFILSTH-----YNLHNLGLTEALASHRALVKAR-GTRPPFVISRS 877
DB 569 NOQGDHD-LATHAISPNATHADGTVEYNIYIIGLQERAIYEAALLEIHPNKPFFIIGRS 627
QY 878 TFAGHRYAGHWTDVWSSWEQIASSVEILOPNLLGVPLVGDVCGFLGNTSEELCVRW 937
DB 628 SFAGSKYGMHWSNDYADYMYFISIPQALSMLSGIPFPGVDACGFGNGTDMELCSRW 687
QY 938 TOLGAYPPWRNHNLSLLOEYSSSEPAQAQWRKALTRYALLPHLYTLFHOAHVAGE 997
DB 688 MOLASFPFRRNHNVLGAPOEYVW-EAMKATKTSINRYSLPYYTYLLHESHVTGI 746
QY 998 TVARPLFLFPKDSSTWTDHQLLWGEALLITPVLOAGKAEVGVYFP--LGTWYDLOQTVR 1055
DB 747 PINRAFNWQFPYSKELAGVDTQFVGDAALLVTVLEPGVNHVTKGIFPGENAVYDYFT-- 804
QY 1056 IEALGSLPPPPAAPREPAIHSEGOVWTLPAPLDTINVHLRAGYIIPLOQPGLTSTESROQ 1115
DB 805 -----HKQKFTAGKNETLAAPLGHIPLHIKGGNIIPQEPGYTTTESRKN 850
QY 1116 PMALVALTKGEGARGELFWDDGESLEVLERGAYTOVIFLARNTIWNELVRYTSEGAGL 1175
DB 851 PFGLLVALDABGTASGLYLDGSEVDV-EALY--VDFVASKNKL--ASVFGSEYVR 904
QY 1176 Q-LOKVTVLGATAPQOVLNSGVVSNFTYSPDTKVLDDICVSLLMG---EOFLVSW 1227
DB 905 OPLANVTILOVDSEPKKVLFNNETVSHVYENGAVYLTDLKFTKEGAFABEFSIQW 960

RESULT 10
US-09-744-926-2
; Sequence 2, Application US/09744926
; Patent No. 679458
; GENERAL INFORMATION:
; APPLICANT: FROBERG, Claus
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING AN ALPHA-GLUCOSIDASE, PLANTS WHICH
; TITLE OF INVENTION: SYNTHESIZE A MODIFIED STARCH, THE GENERATION OF THE PLANTS, THEIR
; TITLE OF INVENTION: MODIFIED STARCH
; FILE REFERENCE: 514413-3865
; CURRENT APPLICATION NUMBER: US/09/744,926
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/EP99/05536
; PRIOR FILING DATE: 1999-07-31
; PRIOR APPLICATION NUMBER: DE 198 36 097.5
; PRIOR FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-744-926-2

Query Match 20.2%; Score 1314.5; DB 4; Length 682;
Best Local Similarity 40.0%; Pred. No. 5.3e-105;
Matches 279; Conservative 118; Mismatches 221; Indels 79; Gaps 19;

QY 571 HPFYL-----ALEDG-----GSAHGVFLNSNAMDVLQSPALSWRSTGGILDVYIFL 619
DB 13 HPQLHRPALHRYGYSFRYPAGVSHGVLLSSNGMDIV-YTGDRISKYKIVGLIDLYFFA 71
QY 620 GREPKSVQOYLDVGVYPPMPYWGILGFHLCRWYSGSTAITQVVENMTRAFPLDVQVN 679
DB 72 GPSPEMVVDQYTLQIRPAAMPYWSFGHCRWGYKNIDDELVDVDSYAKSRIPLEVMWT 131
QY 680 DLDYMSRDRDFTNKGDFDFA-----MVQELHQGRRYMMIVDPAISSGPAGSYRYV 734
DB 132 DIDYMGDFKDFLDP---VNFPLERVIFFLKRLHQNDQKTVLVDPGISIN---NTDYT 185

QY 735 DEGLRGVPIETNETGOPLIGKWPGSTAPDPTNPTALAWEDMVAEFHDOVPPDGLWID 794
DB 186 RRGMEADVEIKRD-NMPYQGVVMPGNVYYPDLINPATEVFWRNIEIKFQDLVFFDGLWLD 244
QY 795 MNEPNSFIRGSDGCPNNELENPPYVGVVGGTL--QAATICASSHQFLST-HYNLHNL 851
DB 245 MNELSNFI--TSPPPTFSSTFDDPPYKINNSGDHLPINRYTVPATSTHFGDTMEYNVHLY 302
QY 852 GLTEAIAASHRALVKARGTRPPFVISRSTFAGHRYAGHWTDVWSSWEQIASSVPEILOPN 911
DB 303 GLEGRATYSALVNVTKRPFILVRSTPLGSKRYTSHMTGDNAATWNDLAYSIPITLSFG 362
QY 912 ILGVPLVGDVCGFLGNTSEELCVRWTOIGAYPPWRNHNLSLLOEYSSSEPAQAQW 971
DB 363 LFGIPVMVGADICGFSNTTEELCRRWIQLGAYPPFARDHSAKDTTPOEILYSWDSVA-AA 421
QY 972 RKALTRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTDHQLLWGEALLITPV 1031
DB 422 KKVGLGRYQLLYPFYMLAYEAHKGTPIARPLFPFPQDAKTFDSTQFLGKGVWISPI 481
QY 1032 LOAGKAEVGVYPLGTWYDL---QTVPIEALGSLPPPPAAPREPAIHSEGOVWTLPAPL 1087
DB 482 LKQGATSDVAYFPAGNWFDFNYSRSVL-----NOGTYMTLDAPP 522
QY 1088 DTINVHLRAGYIIPLOQPGLTSTESROQPMALVALTKGEGARGELFWDDGESLEV-LER 1146
DB 523 DHINHVREGNTLVNQGEAMTTQAAQRTAFKLLVLLSSKNSTGELFVDDDDDEVQMGREG 582
QY 1147 GAYTOVIFLARNTIWNELVRYTSE-----GAGLOLQKVTVLG-----VAT 1187
DB 583 GRWTLVKP---NSNIIGNKIVKSEVNGRYALDQGLVLEKVTLLGFGVNRGLKSYELVG 639
QY 1188 APQOVLNSGVVSNFTYSPDTKVLDDIC-VSLLMGSEQF 1223
DB 640 SHQQ--GNTMKESLKQSCQFVTMEISGMSILIGKEF 674

RESULT 11
US-09-949-016-10627
; Sequence 10627, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10627
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10627

Query Match 16.6%; Score 1083.5; DB 4; Length 1003;
Best Local Similarity 44.7%; Pred. No. 1.4e-84;
Matches 219; Conservative 79; Mismatches 149; Indels 43; Gaps 10;

QY 756 VMPGSTAPDPTNPTALAWEDMVAEFHDOVPPDGLWIDMNEPNSFIRGSDGCPNNELE 815
DB 1 VMPGLTVYPDFTNPCIDWANECSIFHQEVQYDGLWIDMNEVSSFIQSTGKGVNKLN 60
QY 816 NPYPYGVVGGTLOQAATTCASSHQFLSTHYNLHNLGLTEAIAASHRALVKA-RGTRPPVI 874
DB 61 YPFPTDILDKLWYSKTIKCMQAVQNGKQYDVHSLYGSMAIATEQAVQKVPFNKRKSFIL 120


```

Db      61  LSDTKGVLHSRHTLKEKLSRINQGLDRLRRVSHQGVSTAEFPEPRVIDLWDLAQSAN 120
Qy      255  LTDKELEAFREELKHFEAKIEKHNHYQKQLEIAHEKLRHAESVGDGERVSRREKHALL 314
Db      121  LTDKELEAFREELKHFEAKIEKHNHYQKQLEIAHEKLRHAESVGDGERVSRREKHALL 180
Qy      315  GRTELGYTVVKKHLODLSGRISRAR 339
Db      181  GRTELGYTVVKKHLODLSGRISRAR 205

RESULT 14
US-08-216-593-2
; Sequence 2, Application US/08216593
; Patent No. 5650391
; GENERAL INFORMATION:
; APPLICANT: Broze, George J., Jr.
; APPLICANT: Schwartz, Alan L.
; APPLICANT: Warshawsky, Ilka
; TITLE OF INVENTION: Methods and Compositions for Inhibition
; TITLE OF INVENTION: of Hepatic Clearance of Tissue Factor Pathway Inhibitor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; ADDRESSEE: Corporate Patent Dept.
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,593
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Dennis A.
; REGISTRATION NUMBER: 34,547
; REFERENCE/DOCKET NUMBER: C-2795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (708)470-6501
; TELEFAX: (708)470-6881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-593-2

Query Match      16.2%; Score 1054; DB 1; Length 209;
Best Local Similarity 99.5%; Pred. No. 3e-83;
Matches 204; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      135  PRLEKLWHKAKTSKFGSEELDKLREFLHHKEKVHEYNVLLETLSRTEETHENVISPSD 194
Db      1  PRLEKLWHKAKTSKFGSEELDKLREFLHHKEKVHEYNVLLETLSRTEETHENVISPSD 60
Qy      195  LSDTKGVLHSRHTLKEKLSRINQGLDRLRRVSHQGVSTAEFPEPRVIDLWDLAQSAN 254
Db      61  LSDTKGVLHSRHTLKEKLSRINQGLDRLRRVSHQGVSTAEFPEPRVIDLWDLAQSAN 120
Qy      255  LTDKELEAFREELKHFEAKIEKHNHYQKQLEIAHEKLRHAESVGDGERVSRREKHALL 314
Db      121  LTDKELEAFREELKHFEAKIEKHNHYQKQLEIAHEKLRHAESVGDGERVSRREKHALL 180
Qy      315  GRTELGYTVVKKHLODLSGRISRAR 339
Db      181  GRTELGYTVVKKHLODLSGRISRAR 205

```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 5, 2005, 01:05:12 ; Search time 1601 Seconds
(without alignments)
254.339 Million cell updates/sec

Title: US-10-600-862A-7
Perfect score: 6522
Sequence: 1 MRPGSGALWLLALRTVLGS.....KVLDCVSLLMGGEQLVSWC 1228

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1413372 seqs, 331592847 residues

Total number of hits satisfying chosen parameters: 1413372

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6522	100.0	1228	17 US-10-600-862A-7	Sequence 7, Appli
2	6522	100.0	1228	17 US-10-812-849-7	Sequence 7, Appli
3	4744.5	72.7	952	17 US-10-943-893-2	Sequence 2, Appli
4	2024	31.0	707	14 US-10-032-189-125	Sequence 125, App
5	1877.5	28.8	1827	9 US-09-819-247-2	Sequence 2, Appli
6	1877.5	28.8	1827	15 US-10-428-225-2	Sequence 2, Appli
7	1877.5	28.8	1827	16 US-10-751-699-14	Sequence 14, Appli
8	1783	27.3	972	17 US-10-600-862A-9	Sequence 9, Appli
9	1783	27.3	972	17 US-10-812-849-9	Sequence 9, Appli
10	1702	26.1	463	17 US-10-600-862A-11	Sequence 11, Appli
11	1702	26.1	463	17 US-10-812-849-11	Sequence 11, Appli
12	1660.5	25.5	357	17 US-10-600-862A-21	Sequence 21, Appli
13	1660.5	25.5	357	17 US-10-868-577A-45	Sequence 45, Appli

14	1660.5	25.5	357	17	US-10-868-577A-53	Sequence 53, Appli
15	1660.5	25.5	357	17	US-10-812-849-21	Sequence 21, Appli
16	1660.5	25.5	682	17	US-10-741-600-1267	Sequence 1267, Ap
17	1660.5	25.5	682	17	US-10-741-600-1268	Sequence 1268, Ap
18	1650	25.3	320	17	US-10-812-849-1	Sequence 1, Appli
19	1650	25.3	323	9	US-09-792-793A-31	Sequence 31, Appli
20	1650	25.3	323	15	US-10-375-209A-31	Sequence 31, Appli
21	1650	25.3	323	15	US-10-600-862A-1	Sequence 1, Appli
22	1522.5	23.3	902	13	US-10-043-418-4	Sequence 4, Appli
23	1518	23.3	886	16	US-10-437-963-191618	Sequence 191618,
24	1494.5	22.9	903	13	US-10-043-418-3	Sequence 3, Appli
25	1481	22.7	906	15	US-10-424-599-155772	Sequence 155772,
26	1475.5	22.6	932	16	US-10-437-963-155467	Sequence 155467,
27	1458.5	22.4	914	13	US-10-043-418-2	Sequence 2, Appli
28	1437.5	22.0	752	15	US-10-425-114-57817	Sequence 57817, A
29	1403.5	21.5	877	13	US-10-043-418-1	Sequence 1, Appli
30	1396	21.4	909	16	US-10-437-963-151932	Sequence 151932,
31	1353.5	20.8	1075	15	US-10-369-493-3792	Sequence 3792, Ap
32	1327.5	20.4	378	17	US-10-600-862A-22	Sequence 22, Appli
33	1327.5	20.4	378	17	US-10-812-849-22	Sequence 22, Appli
34	1325.5	20.3	357	17	US-10-600-862A-23	Sequence 23, Appli
35	1325.5	20.3	357	17	US-10-812-849-23	Sequence 23, Appli
36	1292.5	19.8	930	15	US-10-424-599-284885	Sequence 284885,
37	1263	18.4	919	15	US-10-369-493-5379	Sequence 5379, Ap
38	1226.5	18.8	304	16	US-10-408-765A-1913	Sequence 1913, Ap
39	1215	18.6	856	15	US-10-369-493-6567	Sequence 6567, Ap
40	1200	18.4	890	15	US-10-369-493-3115	Sequence 3115, Ap
41	1195.5	18.3	348	17	US-10-600-862A-24	Sequence 24, Appli
42	1195.5	18.3	348	17	US-10-812-849-24	Sequence 24, Appli
43	1166.5	17.9	762	15	US-10-369-493-4166	Sequence 4166, Ap
44	1146	17.6	520	15	US-10-425-114-42226	Sequence 42226, A
45	1061	16.3	209	17	US-10-600-862A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-600-862A-7
; Sequence 7, Application US/10600862A
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE
; DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600, 862A
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GAA fusion sequence
US-10-600-862A-7

Query Match	100.0%	Score	6522	DB	17	Length	1228
Best Local Similarity	100.0%	Pred. No.	0				
Matches	1228	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MRPGSGALWLLALRTVLGSYSREKNQPKPSKRSGEFRMEKLNQNLWEKAQRLHPPV	60				
Db	1	MAGPSGALWLLALRTVLGSYSREKNQPKPSKRSGEFRMEKLNQNLWEKAQRLHPPV	60				
Qy	61	RLAELHADLKIQERDELAWKKLDGLDGDGSKEARLIRNLNVLAKYGLDGKKDARQVT	120				
Db	61	RLAELHADLKIQERDELAWKKLDGLDGDGSKEARLIRNLNVLAKYGLDGKKDARQVT	120				
Qy	121	SNSLSGTQBDGLDPRLEKLMWKAKTSKFSGEELDKLWREFLHHKKEVHYNNVLETLIS	180				
Db	121	SNSLSGTQBDGLDPRLEKLMWKAKTSKFSGEELDKLWREFLHHKKEVHYNNVLETLIS	180				

```
181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLRINQGLDRLRRVSHQGYSTAEFEE 240
181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLRINQGLDRLRRVSHQGYSTAEFEE 240
241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
301 ERVSRREKHALLEGRTKELGYTVKXHLQDLSGRISRAAEAEAGHPRPRAVPTQCDV 360
301 ERVSRREKHALLEGRTKELGYTVKXHLQDLSGRISRAAEAEAGHPRPRAVPTQCDV 360
361 PPNRSRDCAPDKAITOECEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
361 PPNRSRDCAPDKAITOECEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
421 EMGYTATLTRTPTTTPFPKDLTLRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSA 480
421 EMGYTATLTRTPTTTPFPKDLTLRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSA 480
481 PSPLYSVEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQLSTLPSQYITGLAEHL 540
481 PSPLYSVEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQLSTLPSQYITGLAEHL 540
541 SPLMLSTWTRITLWNRDLAPTGANLYGSHHPFLYALEDDGSAHGVFLNLSNAMDVLP 600
541 SPLMLSTWTRITLWNRDLAPTGANLYGSHHPFLYALEDDGSAHGVFLNLSNAMDVLP 600
601 SPALSWRSTGGILDVYIFLGPBKSVVQYLDVGVYPPMPYVGLGFHLCRWGYSSTAIT 660
601 SPALSWRSTGGILDVYIFLGPBKSVVQYLDVGVYPPMPYVGLGFHLCRWGYSSTAIT 660
661 ROVENNMTAHPFLDVQWMDLDYMSRRDFTFNKDGFRDPFPAVQELHGGRRYMMIVDP 720
661 ROVENNMTAHPFLDVQWMDLDYMSRRDFTFNKDGFRDPFPAVQELHGGRRYMMIVDP 720
721 A1SSSGPAGSYRYPDEGLRGVPIINETGQPLIGKWPGSTAFPDPFTNPALAWEDMVA 780
721 A1SSSGPAGSYRYPDEGLRGVPIINETGQPLIGKWPGSTAFPDPFTNPALAWEDMVA 780
781 EFHDQVPPFGLWIDMNEPNFIRGSDGCPNNELENPPYVGVGGLTQAATICASSHOF 840
781 EFHDQVPPFGLWIDMNEPNFIRGSDGCPNNELENPPYVGVGGLTQAATICASSHOF 840
841 LSTHYNLNLYGLTEAIAASHRALVKARGTRPFVISRSTFAGHRYAGHWTGDVWSSWEOL 900
841 LSTHYNLNLYGLTEAIAASHRALVKARGTRPFVISRSTFAGHRYAGHWTGDVWSSWEOL 900
901 ASSVPEILOFNLLGVLGVADVCGFLGNTSEELCVRWTLQGFAPYPMRNHNSLLSLPQP 960
901 ASSVPEILOFNLLGVLGVADVCGFLGNTSEELCVRWTLQGFAPYPMRNHNSLLSLPQP 960
961 YSFSEPAQAMRKALTLRVALLPHLYTLFHOAHVAGETVARPLFLFPFKDSSTWVDHOL 1020
961 YSFSEPAQAMRKALTLRVALLPHLYTLFHOAHVAGETVARPLFLFPFKDSSTWVDHOL 1020
1021 LMGEALLITPVLOAGKAEVTGYPPLGTWYDLQTVPIEALGSLPPPPAAAPREPAIHSEGO 1080
1021 LMGEALLITPVLOAGKAEVTGYPPLGTWYDLQTVPIEALGSLPPPPAAAPREPAIHSEGO 1080
1081 VTLPAPDLTINVHLRAGYIIPLOGPGLTTESTRQCPMALAVALTKEGARGELFWDGDES 1140
1081 VTLPAPDLTINVHLRAGYIIPLOGPGLTTESTRQCPMALAVALTKEGARGELFWDGDES 1140
1141 LEVLERGATQVIFLARNTIIVNELVRVTSSEGAGLOLQKTVLVGATAPQOVLNSGVPS 1200
1141 LEVLERGATQVIFLARNTIIVNELVRVTSSEGAGLOLQKTVLVGATAPQOVLNSGVPS 1200
1201 NFTYSPDTKVLDICVSLMGEQPLVSWC 1228
1201 NFTYSPDTKVLDICVSLMGEQPLVSWC 1228
```

```
RESULT 2
US-10-812-849-7
; Sequence 7, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND
; TITLE OF INVENTION: OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; PRIOR FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GAA fusion sequence
US-10-812-849-7
```

Query Match 100.0%; Score 6522; DB 17; Length 1228;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MRGPGALWLLALRTVLGYSREKNQPKSPKRESGEFRMEKLNQWKAORLHLPV 60
DB 1 MRGPGALWLLALRTVLGYSREKNQPKSPKRESGEFRMEKLNQWKAORLHLPV 60
QY 61 RLAEHLADLKIQERDELAWKKLKLDELDEGDEKEARLIRNLNVLAKYGLDGKKDARQVT 120
DB 61 RLAEHLADLKIQERDELAWKKLKLDELDEGDEKEARLIRNLNVLAKYGLDGKKDARQVT 120
QY 121 SNSLSGTQEDGLDDPRLEKLMHAKTSGKFSGEELDKLWREFLHHKKEVHEYNVLLETLS 180
DB 121 SNSLSGTQEDGLDDPRLEKLMHAKTSGKFSGEELDKLWREFLHHKKEVHEYNVLLETLS 180
QY 181 RTEETHENVISPSDLSDIKGSLVLSRHTLKEKLRINQGLDRLRRVSHQGYSTAEFEE 240
DB 181 RTEETHENVISPSDLSDIKGSLVLSRHTLKEKLRINQGLDRLRRVSHQGYSTAEFEE 240
QY 241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
DB 241 PRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKHNHYQKLEIAHEKLRHAESVGDG 300
QY 301 ERVSRREKHALLEGRTKELGYTVKXHLQDLSGRISRAAEAEAGHPRPRAVPTQCDV 360
DB 301 ERVSRREKHALLEGRTKELGYTVKXHLQDLSGRISRAAEAEAGHPRPRAVPTQCDV 360
QY 361 PPNRSRDCAPDKAITOECEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
DB 361 PPNRSRDCAPDKAITOECEARGCCYIPAKQGLQGAQMGPWCFFPPSPSYKLENLSS 420
QY 421 EMGYTATLTRTPTTTPFPKDLTLRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSA 480
DB 421 EMGYTATLTRTPTTTPFPKDLTLRLDVMETENRLHFTIKDPANRRYEVPLETPRVHSA 480
QY 481 PSPLYSVEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQLSTLPSQYITGLAEHL 540
DB 481 PSPLYSVEFSEEPFGVIVHRQDGRVLLNTTVAFLFFADQFLQLSTLPSQYITGLAEHL 540
QY 541 SPLMLSTWTRITLWNRDLAPTGANLYGSHHPFLYALEDDGSAHGVFLNLSNAMDVLP 600
DB 541 SPLMLSTWTRITLWNRDLAPTGANLYGSHHPFLYALEDDGSAHGVFLNLSNAMDVLP 600
QY 601 SPALSWRSTGGILDVYIFLGPBKSVVQYLDVGVYPPMPYVGLGFHLCRWGYSSTAIT 660
DB 601 SPALSWRSTGGILDVYIFLGPBKSVVQYLDVGVYPPMPYVGLGFHLCRWGYSSTAIT 660
QY 661 ROVENNMTAHPFLDVQWMDLDYMSRRDFTFNKDGFRDPFPAVQELHGGRRYMMIVDP 720
```

Db 661 RQVENWTRAHFPLDVQMDLNDYMSRRDFTFNKDGFRDPPAMVQELHOGGRRYMI VDP 720
Qy 721 A1SSSGPAGSYRYDEGLRGVITNETGOLPIKGVWPGSTAFDPDTNPTALAWEDMVA 780
Db 721 A1SSSGPAGSYRYDEGLRGVITNETGOLPIKGVWPGSTAFDPDTNPTALAWEDMVA 780
Qy 781 EFHQDQVFFDGLWIDMNEPSNFIRGSDGCPNNELENPPYVGVGGTLOAATTCASSHQF 840
Db 781 EFHQDQVFFDGLWIDMNEPSNFIRGSDGCPNNELENPPYVGVGGTLOAATTCASSHQF 840
Qy 841 LSTHYNLHNLGYLTAASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWEQL 900
Db 841 LSTHYNLHNLGYLTAASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWEQL 900
Qy 901 ASSVPEILQFNLLGVLGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSPQEP 960
Db 901 ASSVPEILQFNLLGVLGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSPQEP 960
Qy 961 YSPSEPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQL 1020
Db 961 YSPSEPAQAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQL 1020
Qy 1021 LAGEALLITPVLOAGKAETGYEPLGTWYDLQVPIEALGSLPPPAPAPREPAIHSEGOW 1080
Db 1021 LAGEALLITPVLOAGKAETGYEPLGTWYDLQVPIEALGSLPPPAPAPREPAIHSEGOW 1080
Qy 1081 VTLPAPLDTINVLHLAGYIIPLOGPGLTTTESRQQPMALAVALTCKGEARGELFWDGGS 1140
Db 1081 VTLPAPLDTINVLHLAGYIIPLOGPGLTTTESRQQPMALAVALTCKGEARGELFWDGGS 1140
Qy 1141 LEVLERGAYTQVIFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPS 1200
Db 1141 LEVLERGAYTQVIFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPS 1200
Qy 1201 NFTYSPDTKVLDICVSLLMGEQFLVSWC 1228
Db 1201 NFTYSPDTKVLDICVSLLMGEQFLVSWC 1228

RESULT 3
US-10-943-893-2
; Sequence 2, Application US/10943893
; Publication No. US20050058634A1
; GENERAL INFORMATION:
; APPLICANT: Genzyme Corporation
; TITLE OF INVENTION: Methods for Introducing Mannose 6-Phosphate and Other Oligosaccharides into Cells
; FILE OF INVENTION: onto Glycoproteins and its Application Thereof
; FILE REFERENCE: 07680.0031-01000
; CURRENT APPLICATION NUMBER: US/10/943,893
; CURRENT FILING DATE: 2004-09-20
; PRIOR APPLICATION NUMBER: US 10/051,711
; PRIOR FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 60/263,078
; PRIOR FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 952
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: recombinant human acid alpha glucosidase
; FEATURE:
; NAME/KEY: recombinant human acid alpha glucosidase
; LOCATION: (1)..(952)
; OTHER INFORMATION:
US-10-943-893-2

Query Match 72.7%; Score 4744.5; DB 17; Length 952;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

RESULT 4
US-10-032-189-125
; Sequence 125, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T

Qy 313 LBGRTKELGYTVKKHLQDLISGRISRARAEATGAHPGRPRAPVPTQCDVPNSRFDCAPOK 372
Db 42 LSGSSPVLBEETHPAHQOGASRPGPR-----DAQAHEGRPRAPVPTQCDVPNSRFDCAPOK 96
Qy 373 AITQEOCEARGCCYIPAKOGLQAGQWPCWPPPPSYPSYKLENLSSSMGYATATLTRTT 432
Db 97 AITQEOCEARGCCYIPAKOGLQAGQWPCWPPPPSYPSYKLENLSSSMGYATATLTRTT 156
Qy 433 PTFFPKDILTLRLDVMMETENLHFTIKDIPANRRYEVPLETIPRVHSRAPSPLYSVFSEE 492
Db 157 PTFFPKDILTLRLDVMMETENLHFTIKDIPANRRYEVPLETIPRVHSRAPSPLYSVFSEE 216
Qy 493 PFGVIVHROLDGRVLLNTTVAFLPADQFLQSLPSQYITGLAEHLSPMLSTSWTRI 552
Db 217 PFGVIVHROLDGRVLLNTTVAFLPADQFLQSLPSQYITGLAEHLSPMLSTSWTRI 276
Qy 553 TLWNRDLATPTCANLYGSHPPFYLALEDGGSAGVFLNNSNMDVVLQSPALSWRSTGGI 612
Db 277 TLWNRDLATPTCANLYGSHPPFYLALEDGGSAGVFLNNSNMDVVLQSPALSWRSTGGI 336
Qy 613 LDVYIFLGPPEKSVVQQYLDVVGYPPMPYWGFLGFLCRWYSSSTAITRQVVENMTRAHF 672
Db 337 LDVYIFLGPPEKSVVQQYLDVVGYPPMPYWGFLGFLCRWYSSSTAITRQVVENMTRAHF 396
Qy 673 PLDVQWNLDDYMSRRDFTFNKDGFRDPPAMVQELHOGGRRYMI VDPALSSSGPAGSYR 732
Db 397 PLDVQWNLDDYMSRRDFTFNKDGFRDPPAMVQELHOGGRRYMI VDPALSSSGPAGSYR 456
Qy 733 PYDEGLRRGVITNETGOLPIKGVWPGSTAFDPDTNPTALAWEDMVAHFDOVPDGLW 792
Db 457 PYDEGLRRGVITNETGOLPIKGVWPGSTAFDPDTNPTALAWEDMVAHFDOVPDGLW 516
Qy 793 IDWNEPSNFI RSGSDGCPNNELENPPYVGVGGTLOAATTCASSHQFLSTHYNLHNLGY 852
Db 517 IDWNEPSNFI RSGSDGCPNNELENPPYVGVGGTLOAATTCASSHQFLSTHYNLHNLGY 576
Qy 853 LTEATASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWEQALASSVPEILQFNL 912
Db 577 LTEATASHRALVKARGTRPFVSRSTFAGHGRYAGHWTGDVWSSWEQALASSVPEILQFNL 636
Qy 913 LGVPLVGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSPQESFSEPAQAMR 972
Db 637 LGVPLVGADVCGFLGNTSEELCVRWTLQGFYPPMRNHNLSLLSPQESFSEPAQAMR 696
Qy 973 KALTLYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQLWGEALLITPVL 1032
Db 697 KALTLYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQLWGEALLITPVL 756
Qy 1033 QAGKAETGYFPPLGTWYDLQVPIEALGSLPPPAPAPREPAIHSEGOQWTLPAPLDTINV 1092
Db 757 QAGKAETGYFPPLGTWYDLQVPIEALGSLPPPAPAPREPAIHSEGOQWTLPAPLDTINV 816
Qy 1093 HLRAGYIIPLOGPGLTTTESRQQPMALAVALTCKGEARGELFWDGGSLEVLERGAYTQV 1152
Db 817 HLRAGYIIPLOGPGLTTTESRQQPMALAVALTCKGEARGELFWDGGSLEVLERGAYTQV 876
Qy 1153 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPSNFTYSPDTKVLD 1212
Db 877 IFLARNNTIVNELVRVTSEGAGLQKQVTLGVATAPQOVLNSGVVPSNFTYSPDTKVLD 936
Qy 1213 ICVSLLMGEQFLVSWC 1228
Db 937 ICVSLLMGEQFLVSWC 952

QY 1102 LQPGGLTTTSSRQOPMALAVALTGKGARGELFWDDGESLEVLERGAYTQVIFLARNNTI 1161
Db 799 IQEPDVTTTASRKNPLGLIIVALGENTAKGDFWDDGETKDTIQNGNYLYLTFVSNNTL 858
QY 1162 VNELVRVTS---EGAGLQLOKVTVLGV--ATAPOQVLSNGVPV---SNFTYSPDKVLDI 1213
Db 859 --DIVCTHSSYQEGTTLAFQTVKILGLTDSVTEVRVAENNOPMNAHSNFTYDASNOVLII 916
QY 1214 C-VSLLMGGEQFLVSW 1227
Db 917 ADLKNLGRNFSVQW 931

RESULT 7
US-10-751-699-14
; Sequence 14, Application US/10751699
; Publication No. US20040142389A1
; GENERAL INFORMATION:
; APPLICANT: O'Mahony, Daniel J
; TITLE OF INVENTION: Antibodies To Peptides That Target GIT Receptors And Related Meth
; FILE REFERENCE: P26.477-B USA
; CURRENT APPLICATION NUMBER: US/10/751,699
; PRIOR FILING DATE: 2004-01-05
; PRIOR APPLICATION NUMBER: US 09/443,780
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/109,036
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: hSI receptor
US-10-751-699-14

Query Match 28.8%; Score 1877.5; DB 16; Length 1827;
Best Local Similarity 41.7%; Pred. No. 4e-137;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

QY 340 ABATGAHGRPRAVPTQCDVPNSRFDCAPKAITOEOCEARGCCYIPAKOGLQAQMG 399
Db 49 ATTRVTNPSDGGKCNVLNDPNNVRINCIPQFTEGICAGRGCCWRPNWDSL----- 102
QY 400 QWCEPPPPSPYKLENLSSESGYATLTR--TTPTFFPKDILTLRLDVMWMTENRLHFT 458
Db 103 IPWCFFVDNH-GYNVODMTTTSIGVEAKLNRIPTPLFGNDINSVLFTTQNTPNRFRK 161
QY 459 IKDPANRRYEVEPLETRVHS--RAPSLYSVBFSEPPFGVIVHRQDGRVLNMTTVPAPLF 516
Db 162 ITDPNNRRYEVEHQYKEFTGPTVSDTLDVKAQNPFSIQVIRKSNKGLTFTDSIGPLV 221
QY 517 FADQFLQSTSLPSQVITGLAEHL-SPLMLSTSWRITLWNRDLAP-TPCANLYGSHPPY 574
Db 222 YSDQYLQISARLPDSYIYGISQVHRFRHLSWKTPIFTRDQLPGDNNNNLYGHQTFE 281
QY 575 LALED-GGSAHGVLFLNSAMDVLPSPALSWRSRSGTLDVYVIFLGPPEKSVQOQLDV 633
Db 282 MCIEDTSGKSGFVFLMSNAMEIFQPTPIVIRVTGGILDYFILLIGDTPQVQOYQQL 341
QY 634 VGYPFPYWGGLGFHLICRGYSSTATTRQVENMTTRAHPPLDVQMNLDLMDSRDFTFN 693
Db 342 VGLPAMPAYNLGFLSRNYSKSLDVVKEVVRNREAGIPFTQVTDIDYMEDKDFTYD 401
QY 694 KQGRFPDPAWQELHOGGRYRMVDPDAISSGPGAG--SYRYPDEGLRGVFTINETGO- 750
Db 402 QVAFNLGPQFQDLHDHGQYVILDPALISIGRRANGTYYATYERGNTQHVWINESDGT 461
QY 751 PLIGKVPWGSTAFPPDFTNLTALAWMEDVAEFDQVPPDGLWIDMNEPSNFTIRGSDGCP 810
Db 462 PIIGVEWPGLVTPFTNPNCDWMANECSIFHOEVQYDGLWIDMNEVSSFIQGSTKGCN 521

QY 811 NNLENPYPVPGVGGTTLQAATTCASSHQFLSTHYNLHNLNYGLTEAIAASHRALVKA-RGT 869
Db 522 VNKLNPYPPTDILDKLMYSKTCMDAVQNMCKQYDVHSLYGYSMATATEQAVQKVPFNK 581
QY 870 RPFVLSRSTFAGHGRYAGHTGDDWSSWEOLASSVPEILOFNLLGVPLVYADVCGLGNT 929
Db 582 RSFILTRSTFAGSGRHAHAWLGDNTASWEQMEWSITGMLEFSLFGIPLVADICGFVAET 641
QY 930 SEELCVRTQLCAGFYPPFMRNHNLSLSLPOEPYSFSEPA-----QQAMKALTLRYAL 981
Db 642 TEELCRNMQLGAFYPPFSRNHNS-----DGYEHODPAFFGQNSLLVKSRYLIRYTL 695
QY 982 LPHLYTLFHOAHVAGETVARPLFLFPKDDSTWTVDHQLLWGEALLITPVLOAGKAEVTG 1041
Db 696 LPFLYTLFVKAHVFGETVARPVLHEFYEDTNSWIEDTEFLWGPALLITPVLKQAGADTVSA 755
QY 1042 YPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGQWVTLPAPLDITNVHLRAGYIIP 1101
Db 756 YIPDAIWDYDES-----GAKRP-----WRKQVDMYLPADKIGLHRLGGYIIP 798
QY 1102 LQPGGLTTTSSRQOPMALAVALTGKGARGELFWDDGESLEVLERGAYTQVIFLARNNTI 1161
Db 799 IQEPDVTTTASRKNPLGLIIVALGENTAKGDFWDDGETKDTIQNGNYLYLTFVSNNTL 858
QY 1162 VNELVRVTS---EGAGLQLOKVTVLGV--ATAPOQVLSNGVPV---SNFTYSPDKVLDI 1213
Db 859 --DIVCTHSSYQEGTTLAFQTVKILGLTDSVTEVRVAENNOPMNAHSNFTYDASNOVLII 916
QY 1214 C-VSLLMGGEQFLVSW 1227
Db 917 ADLKNLGRNFSVQW 931

RESULT 8
US-10-600-862A-9
; Sequence 9, Application US/10600862A
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE
; TITLE OF INVENTION: DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600,862A
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-1DU fusion sequence
US-10-600-862A-9

Query Match 27.3%; Score 1783; DB 17; Length 972;
Best Local Similarity 99.4%; Pred. No. 3.6e-130;
Matches 346; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRGPSALWLLALRTVLGYSREKNQPKSPKRESGEEFRMEKLNQWKAORLHLPV 60
Db 1 MRGPSALWLLALRTVLGYSREKNQPKSPKRESGEEFRMEKLNQWKAORLHLPV 60
QY 61 RLAEHLADLKIQERDELAWKKLKLDDGDEGEKEARLIRNLNVLAKYGLDGKDKARQVT 120
Db 61 RLAEHLADLKIQERDELAWKKLKLDDGDEGEKEARLIRNLNVLAKYGLDGKDKARQVT 120
QY 121 SNSLSGTQEDGLDDPRLEKLMHAKTSKFGSEEDKLWREPLHHKKEVHENVLLETIS 180
Db 121 SNSLSGTQEDGLDDPRLEKLMHAKTSKFGSEEDKLWREPLHHKKEVHENVLLETIS 180
QY 181 RTEETHENVSPDLSDKGSLHSRHTLEKLSINQGLDRLRVSHOGYSTAEPEE 240

Db 181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLSINQGLDRLRLRVSHQGYSSTAEFEE 240
Qy 241 PRVIDLWDLAQSANLTKELEAFREELKHFEAKIEKHNYQKQLETAHEKLRHAESVGDG 300
Db 241 PRVIDLWDLAQSANLTKELEAFREELKHFEAKIEKHNYQKQLETAHEKLRHAESVGDG 300
Qy 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348
Db 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348

RESULT 9
US-10-812-849-9
; Sequence 9, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; PRIOR FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-IDU fusion sequence
US-10-812-849-9

Query Match 27.3%; Score 1783; DB 17; Length 972;
Best Local Similarity 99.4%; Pred. No. 3.6e-130; Indels 0; Gaps 0;
Matches 346; Conservative 0; Mismatches 130

Qy 1 MRGPSALWLLALRTVLGYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPV 60
Db 1 MRGPSALWLLALRTVLGYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPV 60
Qy 61 RLAEHLADLKIQRDELAWKLLDGLDDEGEKEARLIRNLNVLAKYGLDGKKDARQVT 120
Db 61 RLAEHLADLKIQRDELAWKLLDGLDDEGEKEARLIRNLNVLAKYGLDGKKDARQVT 120
Qy 121 SNSLSGTQEDGLDPPLEKLWHKAKTSGKPSGEBELDKLWREFLHHKEKVEHYNNVLTLS 180
Db 121 SNSLSGTQEDGLDPPLEKLWHKAKTSGKPSGEBELDKLWREFLHHKEKVEHYNNVLTLS 180
Qy 181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLSINQGLDRLRLRVSHQGYSSTAEFEE 240
Db 181 RTEEIHENVISPSDLSDIKGSLVLSRHTLKEKLSINQGLDRLRLRVSHQGYSSTAEFEE 240
Qy 241 PRVIDLWDLAQSANLTKELEAFREELKHFEAKIEKHNYQKQLETAHEKLRHAESVGDG 300
Db 241 PRVIDLWDLAQSANLTKELEAFREELKHFEAKIEKHNYQKQLETAHEKLRHAESVGDG 300
Qy 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348
Db 301 ERVSRREKHALLGRTKELGYTVKXHLQDLSGRISRARAETGAHP 348

RESULT 10
US-10-600-862A-11
; Sequence 11, Application US/10600862A
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600,862A

; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GDNF fusion sequence
US-10-600-862A-11

Query Match 26.1%; Score 1702; DB 17; Length 463;
Best Local Similarity 90.6%; Pred. No. 2.4e-124; Indels 12; Gaps 2;
Matches 338; Conservative 4; Mismatches 19

Qy 19 GSYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPVRLAEHLADLKIQRDELA 78
Db 3 GSYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPVRLAEHLADLKIQRDELA 62
Qy 79 WKLLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKKDARQVTSNSLSGTQEDGLDPPLE 138
Db 63 WKLLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKKDARQVTSNSLSGTQEDGLDPPLE 122
Qy 139 KLWHKAKTSGKPSGEBELDKLWREFLHHKEKVEHYNNVLTLSRTEEIHENVISPSLSDI 198
Db 123 KLWHKAKTSGKPSGEBELDKLWREFLHHKEKVEHYNNVLTLSRTEEIHENVISPSLSDI 182
Qy 199 KGSVLHSRHTLKEKLSINQGLDRLRLRVSHQGYSSTAEFEPVRVIDLWDLAQSANLTK 258
Db 183 KGSVLHSRHTLKEKLSINQGLDRLRLRVSHQGYSSTAEFEPVRVIDLWDLAQSANLTK 242
Qy 259 ELEAFREELKHFEAKIEKHNYQKQLETAHEKLRHAESVGDGERVSRREKHALLEGRTK 318
Db 243 ELEAFREELKHFEAKIEKHNYQKQLETAHEKLRHAESVGDGERVSRREKHALLEGRTK 302
Qy 319 ELGYTVKXHLQDLSGRISRARAETGAH-----PCRPRAVPTQCDVPPNSRFPDCA 371
Db 303 ELGYTVKXHLQDLSGRISRARAETGSPDKQMAVLPRERNRQAAAAANPENS 357

Qy 372 KAITOQCEARGC 384
Db 358 KGRGQRGKRGCC 370

RESULT 11
US-10-812-849-11
; Sequence 11, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; CURRENT FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: RAP-GDNF fusion sequence
US-10-812-849-11

Query Match 26.1%; Score 1702; DB 17; Length 463;
Best Local Similarity 90.6%; Pred. No. 2.4e-124; Indels 12; Gaps 2;
Matches 338; Conservative 4; Mismatches 19

Qy 19 GSYSREKQPKSPKRESGEFRMEKLNQWKAQRLHLPVRLAEHLADLKIQRDELA 78

```
Db      3  GSYSREKNQPKSPKRESGEEFRMEKLNQWKAQRLHLPVRLAELHADLKIQRDELA 62
QY      79  WKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDKARQVTSNSLSGTQEDGLDPRLE 138
Db      63  WKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDKARQVTSNSLSGTQEDGLDPRLE 122
QY      139  KLWHKAKTSGKFSGEELDKLWREFLHHKEKVHEYNVLLTSLSTEBIHENVISPSDLSI 198
Db      123  KLWHKAKTSGKFSGEELDKLWREFLHHKEKVHEYNVLLTSLSTEBIHENVISPSDLSI 182
QY      199  KGSVLHSRHTELKEKLSRINOGLDRLRRVSHQSYSTEAFEEPRVIDLWDLAQSANLTK 258
Db      183  KGSVLHSRHTELKEKLSRINOGLDRLRRVSHQSYSTEAFEEPRVIDLWDLAQSANLTK 242
QY      259  ELEAFREELKHFAKIEKKNHYQKQLEIAHEKLRHAESVGDGERSRSREKHALLEGRTK 318
Db      243  ELEAFREELKHFAKIEKKNHYQKQLEIAHEKLRHAESVGDGERSRSREKHALLEGRTK 302
QY      319  ELGYTVKKHLQDLSGRISRARAETGAH-----PGRPRAVPTQCDVPPNRSRFDCAPD 371
Db      303  ELGYTVKKHLQDLSGRISRARAETGSPDKQMAVLPFRERNRQAAAAANPENS-----G 357
QY      372  KAITQEQCEARGC 384
Db      358  KGRRGQGRGNRC 370

RESULT 12
US-10-600-862A-21
; Sequence 21, Application US/10600862A
; Publication No. US20050026823A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: USE OF THE CHAPERONE RECEPTOR-ASSOCIATED PROTEIN (RAP) FOR THE
; TITLE OF INVENTION: DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND OTHER TISSUES
; FILE REFERENCE: 30610/39383
; CURRENT APPLICATION NUMBER: US/10/600,862A
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-600-862A-21

Query Match      25.5%; Score 1660.5; DB 17; Length 357;
Best Local Similarity 95.1%; Pred. No. 2.8e-121;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

QY      1  MRGPGALWLLAL-----RTVLGSGYSREKNQPKSPKRESGEEFRMEKLNQWKAQRL 55
Db      10  LRGLPALLLLLLFLGWPAAASHGGKYSREKNQPKSPKRESGEEFRMEKLNQWKAQRL 69
QY      56  HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 115
Db      70  HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 129
QY      116  ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSGKFSGBELDKLWREFLHHKEKVHEYNVL 175
Db      130  ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSGKFSGBELDKLWREFLHHKEKVHEYNVL 189
QY      176  LETLSRTEIHENVISPSDLSIKGSLVLSRHTELKEKLSRINOGLDRLRRVSHQSYSTE 235
Db      190  LETLSRTEIHENVISPSDLSIKGSLVLSRHTELKEKLSRINOGLDRLRRVSHQSYSTE 249
QY      236  AEFEEPRVIDLWDLAQSANLTKLEAFREELKHFAKIEKKNHYQKQLEIAHEKLRHAE 295
Db      250  AEFEEPRVIDLWDLAQSANLTKLEAFREELKHFAKIEKKNHYQKQLEIAHEKLRHAE 309
QY      296  SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339
Db      310  SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

RESULT 13
US-10-868-577A-45
; Sequence 45, Application US/10868577A
; Publication No. US20050032697A1
; GENERAL INFORMATION:
; APPLICANT: Alitalo et al.
; TITLE OF INVENTION: HEPARIN BINDING VEGFR-3 LIGANDS
; FILE REFERENCE: 28967/39359A
; CURRENT APPLICATION NUMBER: US/10/868,577A
; PRIOR FILING DATE: 2004-06-14
; PRIOR APPLICATION NUMBER: US 60/478,390
; PRIOR FILING DATE: 2003-06-12
; PRIOR APPLICATION NUMBER: US 10/669,176
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-868-577A-45

Query Match      25.5%; Score 1660.5; DB 17; Length 357;
Best Local Similarity 95.1%; Pred. No. 2.8e-121;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

QY      1  MRGPGALWLLAL-----RTVLGSGYSREKNQPKSPKRESGEEFRMEKLNQWKAQRL 55
Db      10  LRGLPALLLLLLFLGWPAAASHGGKYSREKNQPKSPKRESGEEFRMEKLNQWKAQRL 69
QY      56  HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 115
Db      70  HLPVRLAELHADLKIQRDELAWKKLKLDGLDDEGEKEARLIRNLNVLAKYGLDGKDK 129
QY      116  ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSGKFSGBELDKLWREFLHHKEKVHEYNVL 175
Db      130  ARQVTSNSLSGTQEDGLDPRLEKLWHKAKTSGKFSGBELDKLWREFLHHKEKVHEYNVL 189
QY      176  LETLSRTEIHENVISPSDLSIKGSLVLSRHTELKEKLSRINOGLDRLRRVSHQSYSTE 235
Db      190  LETLSRTEIHENVISPSDLSIKGSLVLSRHTELKEKLSRINOGLDRLRRVSHQSYSTE 249
QY      236  AEFEEPRVIDLWDLAQSANLTKLEAFREELKHFAKIEKKNHYQKQLEIAHEKLRHAE 295
Db      250  AEFEEPRVIDLWDLAQSANLTKLEAFREELKHFAKIEKKNHYQKQLEIAHEKLRHAE 309
QY      296  SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339
Db      310  SVGDGERSRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

RESULT 14
US-10-868-577A-53
; Sequence 53, Application US/10868577A
; Publication No. US20050032697A1
; GENERAL INFORMATION:
; APPLICANT: Alitalo et al.
; TITLE OF INVENTION: HEPARIN BINDING VEGFR-3 LIGANDS
; FILE REFERENCE: 28967/39359A
; CURRENT APPLICATION NUMBER: US/10/868,577A
; PRIOR FILING DATE: 2004-06-14
; PRIOR APPLICATION NUMBER: US 60/478,390
; PRIOR FILING DATE: 2003-06-12
; PRIOR APPLICATION NUMBER: US 10/669,176
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 53
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-10-868-577A-53

Query Match 25.5%; Score 1660.5; DB 17; Length 357;
Best Local Similarity 95.1%; Pred. No. 2.8e-121;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;
QY 1 MRGPGSALWLLAL-----RTVLGSYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 55
DB 10 LRGLPALLLLFLGPWPAASHGCKYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 69
QY 56 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 115
DB 70 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 129
QY 116 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSKFSGEELDKLWREFLHHKEKVHEYNYL 175
DB 130 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSKFSGEELDKLWREFLHHKEKVHEYNYL 189
QY 176 LETLSRTEIHNENIVSPDLSIDIKGVLHSRHTLKEKLSRNOGLDRLRRVSHQGYSTE 235
DB 190 LETLSRTEIHNENIVSPDLSIDIKGVLHSRHTLKEKLSRNOGLDRLRRVSHQGYSTE 249
QY 236 AEFEPFRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKNHNYOKLEIAHEKLRHAE 295
DB 250 AEFEPFRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKNHNYOKLEIAHEKLRHAE 309
QY 296 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339
DB 310 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

RESULT 15

US-10-812-849-21
; Sequence 21, Application US/10812849
; Publication No. US20050042227A1
; GENERAL INFORMATION:
; APPLICANT: Zankel et al.
; TITLE OF INVENTION: MEGALIN-BASED DELIVERY OF THERAPEUTIC COMPOUNDS TO THE BRAIN AND
; FILE REFERENCE: 30610/40037
; CURRENT APPLICATION NUMBER: US/10/812,849
; PRIOR FILING DATE: 2004-03-30
; PRIOR APPLICATION NUMBER: US 10/600,862
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-812-849-21

Query Match 25.5%; Score 1660.5; DB 17; Length 357;
Best Local Similarity 95.1%; Pred. No. 2.8e-121;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;
QY 1 MRGPGSALWLLAL-----RTVLGSYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 55
DB 10 LRGLPALLLLFLGPWPAASHGCKYSREKNQPKSPKRESGEFRMEKLNQLWEKAQRL 69
QY 56 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 115
DB 70 HLPVRLAEHLADLKIQRDELAWKKLKDGLDDEGEKEARLIRNLNVLAKYGLDGKKD 129
QY 116 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSKFSGEELDKLWREFLHHKEKVHEYNYL 175
DB 130 ARQVTSNSLSGTQEDGLDPRLEKLMHKAKTSKFSGEELDKLWREFLHHKEKVHEYNYL 189
QY 176 LETLSRTEIHNENIVSPDLSIDIKGVLHSRHTLKEKLSRNOGLDRLRRVSHQGYSTE 235
DB 190 LETLSRTEIHNENIVSPDLSIDIKGVLHSRHTLKEKLSRNOGLDRLRRVSHQGYSTE 249
QY 236 AEFEPFRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKNHNYOKLEIAHEKLRHAE 295

DB 250 AEFEPFRVIDLWDLAQSANLTDKELEAFREELKHFEAKIEKNHNYOKLEIAHEKLRHAE 309
QY 296 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339
DB 310 SVGDGERVSRSEKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

Search completed: April 5, 2005, 01:43:08
Job time : 1604 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2005, 22:21:01 ; Search time 52 Seconds
(without alignments)
2272.196 Million cell updates/sec

Title: US-10-600-862A-7

Perfect score: 6522
Sequence: 1 MRPGSALWLLALRTVLGS.....KVLDCVSLLMGEQFLVSWC 1228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR 79:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4731.5	72.5	952	1 A32609	alpha-glucosidase
2	1877.5	28.8	1827	1 U0HU	sucrose alpha-gluc
3	1869.5	28.7	1827	1 A23945	sucrose alpha-gluc
4	1802.5	27.6	1841	2 T10799	sucrose alpha-gluc
5	1660.5	25.5	357	2 A39875	alpha-2-macroglobu
6	1522.5	23.3	902	2 T48531	alpha-glucosidase
7	1494.5	22.9	903	2 T09143	alpha-glucosidase
8	1458.5	22.4	913	2 JC5463	alpha-glucosidase
9	1449.5	22.2	915	2 H96709	hypothetical prote
10	1416	21.7	1743	2 T15893	hypothetical prote
11	1403.5	21.5	877	2 S65057	alpha-glucosidase
12	1380	21.2	993	2 T38598	probable family 31
13	1361.5	20.9	995	2 T50267	probable family 31
14	1330.5	20.4	864	2 JC4824	alpha-glucosidase
15	1327.5	20.4	360	2 JX0281	alpha-2-macroglobu
16	1325.5	20.3	357	2 A46646	Heymann nephritis
17	1306	20.0	855	2 T47534	hypothetical prote
18	1263	19.4	919	2 T16693	hypothetical prote
19	1262	19.3	985	2 JC4217	alpha-glucosidase
20	1232.5	18.9	958	1 JN0102	glucan 1,4-alpha-g
21	1215	18.6	856	2 T22575	hypothetical prote
22	1117	17.1	719	2 JC1200	alpha-glucosidase
23	1090	16.7	1070	2 S19686	alpha-glucosidase
24	895	13.7	779	2 AE2402	alpha-glucosidase
25	891.5	13.7	919	2 T07391	probable alpha-glu
26	891.5	13.7	941	2 T32449	hypothetical prote
27	882	13.5	763	2 AH1097	alpha-glucosidase
28	879	13.5	763	2 AG1460	alpha-glucosidase
29	831	12.7	954	2 S46105	glucan 1,4-alpha-g

30	830.5	12.7	818	2 AC2472	alpha-glucosidase
31	779	11.9	910	2 T22050	hypothetical prote
32	779	11.9	924	2 T22044	hypothetical prote
33	759	11.6	642	2 S11386	sucrose alpha-gluc
34	664.5	10.2	693	2 H90486	alpha-glucosidase
35	559.5	8.6	275	2 A36690	sucrose alpha-gluc
36	505.5	7.8	731	2 D90483	alpha-xyloisidase
37	503.5	7.7	801	2 H83737	glucosidase BH0704
38	491	7.5	764	2 A72394	alpha-xyloisidase -
39	490	7.5	792	2 AD0104	probable glucosida
40	454.5	7.0	1090	2 AG1749	glucosidase homolo
41	453.5	7.0	1091	2 AF1380	glucosidase homolo
42	452	6.9	769	2 H97033	alpha-glucosidase
43	435.5	6.7	773	2 AB3888	hypothetical prote
44	431	6.6	679	2 AD0507	probable glycosyl
45	430.5	6.6	657	2 G83906	hypothetical prote

ALIGNMENTS

RESULT 1

A32609

Alpha-glucosidase (EC 3.2.1.20) precursor, lysosomal - human
N:Alternate names: acid alpha-glucosidase; acid maltase; maltase-glucoamylase
C:Species: Homo sapiens (man)
C:Date: 21-May-1990 #sequence revision 05-Apr-1995 #text change 18-Jun-1999
C:Accession: A40577; A32609; A35698; S00831; S18847; I52309; S63526
R:Martiniuk, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.
DNA Cell Biol. 10, 283-292, 1991

A:Title: Isolation and partial characterization of the structural gene for human acid al
A:Reference number: A40577; MUID:91229698; PMID:1674202
A:Accession: A40577

A:Status: not compared with conceptual translation
A:Molecule type: DNA

A:Residues: 1-952 <MAR>

A>Note: order of exons in Figure 2 is incorrect

R:Martiniuk, F.; Mehler, M.; Tzall, S.; Meredith, G.; Hirschhorn, R.
DNA Cell Biol. 9, 85-94, 1990

A:Title: Sequence of the cDNA and 5'-flanking region for human acid alpha-glucosidase, de

s with previous cDNA and amino acid sequences.
A:Reference number: A32609; MUID:90262851; PMID:2111708

A:Accession: A32609

A:Molecule type: mRNA

A:Residues: 1-952 <MA2>

A:Cross-references: GB:M34424; NID:g182907; PIDN:AAA52506.1; PID:g182908

R:Martiniuk, F.; Bodkin, M.; Tzall, S.; Hirschhorn, R.
Am. J. Hum. Genet. 47, 440-445, 1990

A:Title: Identification of the base-pair substitution responsible for a human acid alpha

A:Reference number: A35698; MUID:90365036; PMID:2203258

A:Contents: partial sequence of GAA 2 allelic form

A:Accession: A35698

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-90, 'N', 92-129 <MA3>

R:Hoefsloot, L.H.; Hoogeveen-Westerveld, M.; Kroos, M.A.; van Beeumen, J.; Reuser, A.J.J.
EMBO J. 7, 1697-1704, 1988

A:Title: Primary structure and processing of lysosomal alpha-glucosidase; homology with t

A:Reference number: S00831; MUID:89005058; PMID:3049072

A:Accession: S00831

A:Molecule type: mRNA

A:Residues: 1-156, 'VLL', 130-198, 'R', 200-222, 'H', 224-371, 'L', 373-401, 'R', 403-412, 'TSRSTRM

A:Cross-references: EMBL:X00839

A:Accession: S18847

A:Molecule type: protein

A:Residues: 70-89; 123-126, 'VLL', 130-145; 204-215; 230-249; 332-345; 349-401, 'R', 403-4

R:Lin, C.Y.; Shieh, J.J.
Biochem. Biophys. Res. Commun. 208, 886-893, 1995

A:Title: Identification of a de novo point mutation resulting in infantile form of Pompe'

A:Reference number: I52309; MUID:95209708; PMID:7695647

A:Accession: I52309

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 631-644, 'H', 646-680 <RES>
A:Cross-references: GB:S76893; NID:g912786; PIDN:AAB33842.1; PID:g912787
A>Note: homozygous mutation of Asp-645 to His causes an infantile form of Pompe's disease
R:Fuller, M.; van der Ploeg, A.; Reuser, A.J.J.; Anson, D.S.; Hopwood, J.J.
Eur. J. Biochem. 234, 903-909, 1995
A:Title: Isolation and characterization of a recombinant, precursor form of lysosomal alpha
A:Reference number: S63526; MUID:96163476; PMID:8575451
A:Accession: S63526
A:Molecule type: protein
A:Residues: 68-74, 'X', 76-79, 'X', 81 <PUL>
C:Comment: In common with other lysosomal enzymes, posttranslational processing includes
processing at both amino and carboxyl ends.
C:Genetics:
A:Gene: GDB:GAA
A:Cross-references: GDB:115965; OMIM:232300
A:Map position: 17q23-17q23
A:Introns: 182/3; 231/2; 286/3; 319/1; 398/3; 442/3; 479/3; 517/3; 546/1; 585/2;
A>Note: GAA 1, GAA 2, and GAA 4 are common alleles in the normal population
C:Function:
A:Description: also has EC 3.1.2.3 activity; hydrolyzes alpha 1,4 and alpha 1,6 glycosid
A>Note: deficiency causes glycogen storage disease type II (Pompe's disease)
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology
C:Keywords: glycoprotein; glycosidase; hydrolase; lysosomal storage disease; lysosome; H
F:82-130/Domain: trefoil homology <TRF>
F:208-867/Domain: sucrase/isomaltase homology <SIM>
F:69-70/Cleavage site: Gln-Ala (unidentified proteinase) #status experimental
F:82-109,92-108,103-127/Diulfide bonds: #status predicted
F:122-123/Cleavage site: Met-Cly (unidentified proteinase) #status experimental
F:140,233/Binding site: carbohydate (Asn) (covalent) #status experimental
F:203-204/Cleavage site: Arg-Ala (unidentified proteinase) #status experimental
F:390,470,492,652,825/Binding site: carbohydate (Asn) (covalent) #status predicted
F:518/Active site: Asp #status predicted

Query Match 72.5%; Score 4731.5; DB 1; Length 952;
Best Local Similarity 96.9%; Pred. No. 2.8e-280;
Matches 888; Conservative 3; Mismatches 20; Indels 5; Gaps 1;

QY	313	LEGRTKELGYTVKKHLQDLISGRISARAEATGAHPRPRAVPTQCDVPNPRFDCAPDK 372
DB	42	LSGSSPVLEETHPAHQGASRPGR-----DAQAHGPRPRAVPTQCDVPNPRFDCAPDK 96
QY	373	AITQEQCEARGCCYIPAKOGLQAGQWCPWCPFPSPYKYLENLSSSEMGYTATLTRTT 432
DB	97	AITQEQCEARGCCYIPAKOGLQAGQWCPWCPFPSPYKYLENLSSSEMGYTATLTRTT 156
QY	433	PTFPFKDILTLRDVNMETENRLHFTIKDPAKRREYVPLETRPVHSHRAPSPLYSVEFSEE 492
DB	157	PTFPFKDILTLRDVNMETENRLHFTIKDPAKRREYVPLETRPVHSHRAPSPLYSVEFSEE 216
QY	493	PGFVIVHRQLDGRVLNTTVAFLFFADQFLQLSTSLPSQYITGLAEHLSPMLSTSWTRI 552
DB	217	PGFVIVRQLDGRVLNTTVAFLFFADQFLQLSTSLPSQYITGLAEHLSPMLSTSWTRI 276
QY	553	TLNWRDLAPTPGANLVGSHPFYALBEDGSAGVFLLNSNAMDVLQPSPALSWRSTGIGI 612
DB	277	TLNWRDLAPTPGANLVGSHPFYALBEDGSAGVFLLNSNAMDVLQPSPALSWRSTGIGI 336
QY	613	LDVYIFLGPBKSVQOQLDVVGVPMPPYVGIHGLCRWGSSTAITQVVENMTRAHF 672
DB	337	LDVYIFLGPBKSVQOQLDVVGVPMPPYVGIHGLCRWGSSTAITQVVENMTRAHF 396
QY	673	PLDVQWNLDDYMSRRDFTFNKDGFRDFFPMVQELHQGGRRYMMIVDPAISSSGPAGSYR 732
DB	397	PLDVQWNLDDYMSRRDFTFNKDGFRDFFPMVQELHQGGRRYMMIVDPAISSSGPAGSYR 456
QY	733	PYDEGLRGVFTINETGQPLIGKWPFGSTAFDFTNPTALAWEDMVAFPHQVFPDGLW 792
DB	457	PYDEGLRGVFTINETGQPLIGKWPFGSTAFDFTNPTALAWEDMVAFPHQVFPDGLW 516
QY	793	IDNNEPSNFRGSDCCPNNELENPPYVGVGTLQQAATICASSHQFLSTHYNLNLGY 852
DB	517	IDNNEPSNFRGSDCCPNNELENPPYVGVGTLQQAATICASSHQFLSTHYNLNLGY 576

QY	853	LTEATASHRALVKARGTRFPVISRSTFAGHGRYAGHWTDVWSSWEQCLASSVPEILQFNL 912
DB	577	LTEATASHRALVKARGTRFPVISRSTFAGHGRYAGHWTDVWSSWEQCLASSVPEILQFNL 636
QY	913	LGVPLVGDVCGFLGNTSEELCVRVLTQLGAFYPPFMRNHNLSLSLQEPYSFSEPAQAQMR 972
DB	637	LGVPLVGDVCGFLGNTSEELCVRVLTQLGAFYPPFMRNHNLSLSLQEPYSFSEPAQAQMR 696
QY	973	KALTURYALLPHLYTLPHQAHVAGETVARPLFLBPPKDSSTWTVDHQLLWGBALLITPVL 1032
DB	697	KALTURYALLPHLYTLPHQAHVAGETVARPLFLBPPKDSSTWTVDHQLLWGBALLITPVL 756
QY	1033	QAGKAEVTGYFPPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGGQWVTLPAPLDTINV 1092
DB	757	QAGKAEVTGYFPPLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGGQWVTLPAPLDTINV 816
QY	1093	HLRAGYIIPLOQPGTLTTESRQOPMALAVALTGGEARGELFWDDGESLEVLERGAYTQV 1152
DB	817	HLRAGYIIPLOQPGTLTTESRQOPMALAVALTGGEARGELFWDDGESLEVLERGAYTQV 876
QY	1153	IFLARNTIIVNELVRVTSEAGLQLOKVTVLGVATAPOQVLSNGVPSNFTYSPDTKVLID 1212
DB	877	IFLARNTIIVNELVRVTSEAGLQLOKVTVLGVATAPOQVLSNGVPSNFTYSPDTKVLID 936
QY	1213	ICVSLLMGEQFLVSWC 1228
DB	937	ICVSLLMGEQFLVSWC 952

RESULT 2
UUHU
sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) [validated]
N:Alternate names: limit dextrinase; small intestinal sucrase/isomaltase (SI)
C:Species: Homo sapiens (man)
C:Date: 19-Nov-1988 #sequence revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: S36082; A27326; S24329; A61136
R:Lacasa, M.
submitted to the EMBL Data Library, December 1991
A:Reference number: S36082
A:Accession: S36082
A:Molecule type: mRNA
A:Residues: 1-1827 <LAC>
A:Cross-references: UNIPROT:P14410; EMBL:X63597; NID:g36644; PIDN:CAA45140.1; PID:g36645
R:Green, F.; Edwards, Y.; Hauri, H.P.; Povey, S.; Ho, M.W.; Pinto, M.; Swallow, D.
Gene 57, 101-110, 1987
A:Title: Isolation of a cDNA probe for a human jejunal brush-border hydrolase, sucrase-is
A:Reference number: A27326; MUID:88112852; PMID:2962903
A:Accession: A27326
A:Molecule type: mRNA
A:Residues: 1-661, 'X', 663-678 <GRE>
A:Cross-references: GB:M22616
R:Chantret, I.; Lacasa, M.; Chevallier, G.; Ruf, J.; Islam, I.; Mantei, N.; Edwards, Y.;
Biochem. J. 285, 915-923, 1992
A:Title: Sequence of the complete cDNA and the 5' structure of the human sucrase-isomalt
A:Reference number: S24329; MUID:92359963; PMID:1353958
A:Accession: S24329
A:Molecule type: mRNA
A:Residues: 1-661, 'F', 663-931 <CHA>
A:Cross-references: EMBL:X63597
R:Corvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.
Gastroenterology 101, 618-625, 1991
A:Title: Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human small inte
A:Reference number: A61136; MUID:91317403; PMID:1677636
A:Accession: A61136
A:Molecule type: protein
A:Residues: 2-14, 'F', 16-20; 1008-1015, 'E', 1017-1021, 'TX', 1024 <GOR>
C:Genetics:
A:Gene: GDB:SI
A:Cross-references: GDB:120377; OMIM:222900
A:Map position: 3q25.2-3q26.2
C:Complex: the two product chains remain associated after cleavage
C:Function: <ISM>

QY 1224 LVSWC 1228
Db 934 IWRWC 938

RESULT 5
A:39875
N:Alternate names: alpha-2-macroglobulin receptor precursor - human
C:Species: Homo sapiens (man)
C:Date: 30-Dec-1991 #sequence revision 30-Dec-1991 #text_change 09-Jul-2004
C:Accession: A39875; S55260; S12984
R:Strickland, D.K.; Ashcom, J.D.; Williams, S.; Battey, F.; Behre, E.; McTigue, K.; Batt
J. Biol. Chem. 266, 13364-13369, 1991
A:Title: Primary structure of alpha-2-macroglobulin receptor-associated protein. Human h
A:Reference number: A39875; MUID:91302371; PMID:1712782
A:Accession: A39875
A:Molecule type: mRNA
A:Residues: 1-357 <STR>
A:Cross-references: UNIPROT:P30533; GB:M63959; NID:g177873; PIDN:AAAS1553.1; PID:g177874
R:Bu, G.; Geuze, H.J.; Strous, G.J.; Schwartz, A.L.
EMBO J. 14, 2269-2280, 1995
A:Title: 39 kDa receptor-associated protein is an ER resident protein and molecular chap
A:Reference number: S55260; MUID:95292967; PMID:7774585
A:Accession: S55260
A:Status: preliminary
A:Molecule type: protein
A:Residues: 35-357 <BUG>
R:Kristensen, T.; Moestrup, S.K.; Gliemann, J.; Bendtsen, L.; Sand, O.; Sottrup-Jensen,
FEBS Lett. 276, 151-155, 1990
A:Title: Evidence that the newly cloned low-density-lipoprotein receptor related protein
A:Reference number: S12984; MUID:91092405; PMID:1702392
A:Accession: S12984
A:Status: preliminary
A:Molecule type: protein
A:Residues: 35-44 <KRI>
C:Genetics:
A:Gene: GDB:LRPAP1; A2MRAP
A:Cross-references: GDB:134668; OMIM:104225
A:Map position: 4p16.3-4p16.3
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparin binding
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-357/Product: alpha-2-macroglobulin receptor-associated protein #status predicted <M
F:50.161/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F:268/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 25.5%; Score 1660.5; DB 2; Length 357;
Best Local Similarity 95.1%; Pred. No. 1e-93;
Matches 327; Conservative 2; Mismatches 10; Indels 5; Gaps 1;

QY 1 MRGPGALWLLAL----RTVLGYSREKNQPKSPKRESGEPRMEKLNQLWEKAQRL 55
Db 10 LRGLPALLLLLLFLGPWPAASHGCKYSREKNQPKSPKRESGEPRMEKLNQLWEKAQRL 69
QY 56 HLPVRLAEHLADLKIQRDELAWKKLKLGLDEDEGEKEARLIRNLNVLAKYGLDGKKD 115
Db 70 HLPVRLAEHLADLKIQRDELAWKKLKLGLDEDEGEKEARLIRNLNVLAKYGLDGKKD 129
QY 116 ARQVTNSLSGTQEDGLDDPRLEKLNHAKTSKFSGEELDKLWREFLHHKEKVHYNVL 175
Db 130 ARQVTNSLSGTQEDGLDDPRLEKLNHAKTSKFSGEELDKLWREFLHHKEKVHYNVL 189
QY 176 LETLSRTTEETHENVISPSDLSDTKGSVLHSHRHELEKLSINQGLDRLRRVSHQGYSTE 235
Db 190 LETLSRTTEETHENVISPSDLSDTKGSVLHSHRHELEKLSINQGLDRLRRVSHQGYSTE 249
QY 236 AEFEEPRVIDLWLAQSANLTDKLEAFREELHKEFAKIEKNHYQKLEIAHEKLRHAE 295
Db 250 AEFEEPRVIDLWLAQSANLTDKLEAFREELHKEFAKIEKNHYQKLEIAHEKLRHAE 309
QY 296 SVGDGERSVRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 339
|||||

Db 310 SVGDGERSVRSREKHALLEGRTKELGYTVKKHLQDLSGRISRAR 353

RESULT 6
T48531
alpha-glucosidase 1 - Arabidopsis thaliana
N:Alternate names: protein T22P22.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T48531
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duisterhoef, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24490
A:Accession: T48531
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-902 <BEV>
A:Cross-references: UNIPROT:O9LYF8; EMBL:AL163814
A:Experimental source: cultivar Columbia; BAC clone T22P22
C:Genetics:
A:Map position: 5
A:Introns: 78/2; 313/1; 390/1; 605/3; 747/1
A:Note: T22P22.110
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology

Query Match 23.3%; Score 1522.5; DB 2; Length 902;
Best Local Similarity 38.3%; Pred. No. 1.2e-84;
Matches 338; Conservative 136; Mismatches 301; Indels 107; Gaps 23;

QY 417 LSSSEMGVYATLTRTP-TFFPKDILTLRLDVMETENRHLFTIKDPANRRVEVPLET-- 473
Db 46 VDSNRQVLTAKLDLIKPSVYAPDIKSLNLHVSLETSELRIRITDSSQORWEIP-ETVI 104
QY 474 PRVHSRAP-----SPLYSVEFSEEPFGVIVHRQLDGRVLLNT 511
Db 105 PRAGNHSRPRFSTEEGGNSPENFLADPSSDLVFLHNTTTFGFSVSRSSGDIILFDS 164
QY 512 V-----APLFFADQFLQSTSLPS--QYITGLAEHLSPMLSTSWRITLWNRDI-APT 562
Db 165 PDSDSNTYFIFKQDFLQSSALPENRSLYIGIGHTKSRFLIPGETMTLWNADIGSEN 224
QY 563 PGANLYGSHPPFYAL-----EDGSAICVFLNNSNMDDVLPSPALSWRSTGGILLDVY 616
Db 225 PDVNDLYGSHPPFYMDVRGSKNEEAGTTHGVLLNNGMDVKYE-GHRTITNVIGGIDLY 283
QY 617 IFLGPPEKSVQOYLDDVVGYPMPYVGLGFHLCRWGSYSTATITROVENMTTRAHPFLDV 676
Db 284 VFAGSPENVMNQYTELIGRPAPMPYWSFGHCYKGVKNSDLEYVVDGYAKAGIPELV 343
QY 677 QNNDLDYMDSRDFTFNKDGFRD--FPAMVQELHOGGRRYMMIVDPAISSSGPAGSYRYPY 734
Db 344 MWTDDYMDGYKDFTLDPVNPFPEDKMQSFVDTLKNGKQKYLILDPGI---GVDSSYGTY 400
QY 735 DGLRGVPITNETQPLGKVMRGSTAPDPDTNPTALAWEDMVAEPHDQVPFDCGLWID 794
Db 401 NRGMEADVFI-KRNGEPYLGVMYVPGKVPYFPDFLNPAAATFWSNEIKOMFEIPLDGLWID 459
QY 795 MNEPSNFRGSDGCPNNELENPPVYVGVGG--TLQAATI CASSHQFLS-THYNLHLY 851
Db 460 MNELSNFTSPUSS--GSLDDPPYKINNSGDKRPINNKTVPATSHFGNISYDAHLY 517
QY 852 GITEAIAASHRALVARGTRPFVIRSTFAGHRYAGHWTGDMVSSWEQIASSVPEILOFN 911
Db 518 GLLEAKATHQAVVDITGKRPFILSRSTFVSSGKYTAHMTGDNAAKWEDLAYSIPGLNFG 577
QY 912 ILGVPLVGDVCGFLGNTSEELCVRWOLGAPYPMRNHNSLLSLPOEYSSEPEQAQM 971
Db 578 LFGIPVMVGADICGFGHDTTEELCRWIOGAPYFPAARDHSHSLGTARQELYLWDSVASSA- 636
QY 972 RKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFPPKDSSTWTVDHQLLWGEALLITPV 1031
Db 637 RKVLGLRMKLLPHLYTLMEAHVSGNPILARPLFFSPQDTKYEIDSOFLGKSNVSPA 696
|||||

```
QY 1032 LQAGKAEVTGYPFLGWTYDLQVPIBALGSLPPPPAAPPREPAIHSGQWVTLPAPLDTIN 1091
Db 697 LKQGA VADAYDPAGNWFDFLNFYSF-AVGG-----DSGKHVLDTPADHVN 741

QY 1092 VHLRAGYIIPLOQPGITTTESRQOPMALAVALTKGGEARGELFWDDGESLVLERGAYTQ 1151
Db 742 VHVREGSIVAMQGEALTTRDARTPYQLLVASRLNISEGLFDGLENLWAGGG--- 798

QY 1152 VIPLARNNTIVNELVVTSEGAGLQ-----LQKTVLGVATAPQOVLNSG 1196
Db 799 ----NRDNLVPRCVTKSVLRSEVNVPEVASKMKWSIGKTVFGFENV-ENVKTYE 853

QY 1197 VPVSNFTYGP-----DTKVLDCV---SLLMGEQF 1223
Db 854 VRTSERLRSPRISLIKTVSDNDPRLSVESKLSLLVGKF 895

RESULT 7
T09143
alpha-glucosidase (EC 3.2.1.20) - spinach
C:Species: Spinacia oleracea (spinach)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T09143
R:Sugimoto, M.; Furui, S.; Suzuki, Y.
Plant Mol. Biol. 33, 765-768, 1997
A:Title: Molecular cloning and characterization of a cDNA encoding alpha-glucosidase from
A:Reference number: Z16585; MUID:97238484; PMID:9132069
A:Accession: T09143
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-903 <SUG>
A:Cross-references: UNIPROT:O04893; EMBL:D86624; NID:g2081626; PIDN:BAA19924.1; PID:g208
A:Experimental source: strain Dash
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology
C:Keywords: glycosidase; hydrolase
F:145-799/Domain: sucrase/isomaltase homology <SIM>

Query Match 22.9%; Score 1494.5; DB 2; Length 903;
Best Local Similarity 37.9%; Pred. NO. 5.9e-83;
Matches 340; Conservative 138; Mismatches 296; Indels 123; Gaps 26;

QY 412 YKLENI-----SSSEMGYTA--TLTRTPFPFKDILTLRLDVMETENRHLFIKDPANR 466
Db 40 YKVKSVKVDGTRRSITLALPOLVKNSSVYGP-DIQLLSITASLESNDRLRVITDAKRR 98

QY 467 YEVPLETPRVHSRAPSPLYSVE-----FSEEPFGV 496
Db 99 WEIPDNILRHQPPPPPHSLSLYKLTLSSTTNRKILLSHNSDLTFSLINTTPGF 158

QY 497 IVHRQLDGRVLLNTTAP-----LFFADQFLQLSTSLPS--QYITGLAEHLSPLMLSTS 548
Db 159 TISRSKTHDVLFDATPDTPNPNTFLIFIDQYHLTLTSSLPGTRAHIVGLGHSKPTFQLAH 218

QY 549 WTRITLWNRDL-APTGALNYGHPYLALED---GGSAGHGVFLNNSMDVVLPSPAL 604
Db 219 NOTLTWRAADIPSSNDPNVNLGHPYMDVRSSPVAGSTHGVLLNNSGMDVEYTG-NRI 277

QY 605 SWRSTGILDVIFLGPESKVVQOYLDVVGVYFMPYVGLFHLRCWYSSTAITRQV 664
Db 278 TYKVIIGIDLVFFAGSPGVQVEQTRVIGRAPMPYVAFGQQCRYGHDVYELQSVV 337

QY 665 ENWTRHFLPDVQWNDLDYMSRRDPT-----FNKDGPRDFPAMVQELHQGRRYMYIVD 719
Db 338 AGYAKAKIPIEVMMWTDIDYDAYKDPTLDPVNFPFLDKMKF---VNNLHNGQKYVILD 394

QY 720 PAISSSGPAGSVRYPDEGLRGRVFTNETGQPLGKWFEGSTAFDFTNPTALAWEDMV 779
Db 395 PGISTN---KTYETIRGKMKHDVFL-KRNGKPYLGSWFGVYFPDFLKPSSALTFTWDEI 450

QY 780 AEFHDQVFPDGLWIDMNEISNFT---SSPIPGSTLDNPPYKINNNSGVMLPIINKTIPPTA 836
Db 451 KRFLNLLPVQGLWIDMNEISNFT---SSPIPGSTLDNPPYKINNNSGVMLPIINKTIPPTA 508
```

```
QY 837 SHQFLSTYHNHLNLYGLTEAISHALVAKARTRPFVTSRSTFAGHRYAGHWGTGDVWSS 896
Db 509 MHYGDIPEYNVHNLPGLYLEARVTRAALIKLTKRPFVLRSSTFSGSKYTAHTWGTGNAAT 568

QY 897 NEQLASSVPEILQFNLLGVPLVGADVCFLGNTSEELCVRWTLQGLAFYFPFNNHSLLSL 956
Db 569 WNDLVYSIPSMDFLGFIPMGADICFLGNTTEELCRRWTLQGLAFYFPFSDHSSLSLGT 628

QY 957 POEYPSFSEPAQQAAMRKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFEFFPKDSSWTW 1016
Db 629 QYELRWESVAASA-RKVLGLRYTLPPYILMYEAQLNGIPIARPLFPFSDDDIKTYGI 687

QY 1017 DQQLWGBALLITPVLAGKAEVTGYPFLGWTYDL--QTVPIEALGSLPPPPAAPPREPAI 1074
Db 688 SSQFLLGKGMVSPVLKPGVSVTAYFPRGNWDFLFDYTRSVTA----- 731

QY 1075 HSEGWVTLPAPLDTINVHLRAGYIIPLOGLFTTTESRQOPMALAVALTKGGEARGELF 1134
Db 732 -STGRYVTLSPDPDHINVIQEGNILAMQGMAMTQAARKTPFFHLLVVMSCDGCASFGE 790

QY 1135 WDDGESLEV-LERGAYTQVIFL---ARNNTIVNELVVTSEGAGLQ---LQKTVLGVAT 1187
Db 791 LDDGVEVTMGVNRGKWTFFKFAASAKQTCITSDV-VSGEFVSVQKWVIDKVTILG--- 846

QY 1188 APQQVLSNGVPVSNFTY-----SPDTK-----VLDIC-VSLLMGSOQF 1223
Db 847 ----LRKGTNGYVTVRTGAVTRKDKSKLKSTPDRKGEFIVABISGLNLLGREF 898

RESULT 8
JC5463
alpha-glucosidase (EC 3.2.1.20) - sugar beet
N:Alternate names: alpha-D-glucoside glucosylhydrolase
C:Species: Beta vulgaris var. altissima (sugar beet)
C:Date: 17-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 09-Jul-2004
C:Accession: JC5463; PC4330
R:Matsui, H.; Iwanami, S.; Ito, H.; Mori, H.; Honma, M.; Chiba, S.
BioSci. Biotechnol. Biochem. 61, 875-880, 1997
A:Title: Cloning and sequencing of a cDNA encoding alpha-glucosidase from sugar beet.
A:Reference number: JC5463; MUID:97321863; PMID:9178565
A:Accession: JC5463
A:Molecule type: mRNA
A:Residues: 1-913 <MAT1>
A:Cross-references: UNIPROT:O04931; DDBJ:D89615; NID:g2217947; PIDN:BAA20343.1; PID:g2191
A:Experimental source: seed; cv. NK-152
A:Accession: PC4330
A:Molecule type: protein
A:Residues: 234-261;310-365;507-541;810-840 <MAT2>
A:Experimental source: seed
C:Comment: This enzyme is an exo-glucosylhydrolase that catalyzes the hydrolysis of alpha-gl
C:Superfamily: lysosomal alpha-glucosidase; sucrase/isomaltase homology; trefoil homology
C:Keywords: glycosidase; hydrolase
F:149-803/Domain: sucrase/isomaltase homology <SIM>

Query Match 22.4%; Score 1458.5; DB 2; Length 913;
Best Local Similarity 37.2%; Pred. NO. 9.4e-81;
Matches 330; Conservative 135; Mismatches 313; Indels 109; Gaps 25;

QY 389 AKQGLQGMQPCWCPFPSPSYKLENSSEMGYTATLTFTTTPFKOILTLRLDVM 448
Db 32 SKNDNQGBAIGVY-----QVKNKVDNKGSTKSLTALLQLIRNSPVYGP-DIHFLSFTAS 85

QY 449 METENRLFTIKDPANRRYEVPLET-PR-VHSRAPSPLYSVE----- 488
Db 86 FEEDTLAIRFDANNRWEIPNEVLPRPPPPPSLSSLOHLKPKIPQNPQPTTVLSH 145

QY 489 -----FSEEPGVIVHRQLDGRVLLNTTAP-----LFFADQFLQLSTSLPSQ--- 531
Db 146 PHSDLAFTLFTTTPGFTYRKSHTDVLFDATPIPSNPTTFLIYKDYQLQLSSSLPAQOA 205

QY 532 YITGLAEHLSPLMLSTSWTRITLWNRDLAP--TPGANLVGSHPPFYLALEDG---GSAHGVF 587
Db 206 HLYLGHEHTKPTFLAHNQILTLWNADIASFNRLNIGSHPPFYMDVRSSPMWGSTHGVF 265
```


C:Accession: T15893
R:Geisel, C.
A:Description: The sequence of *C. elegans* cosmid D2096.
A:Reference number: Z18425
A:Accession: T15893
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1743 <GEI>
A:Cross-references: UNIPROT:Q19004; EMBL:U40800; NID:g1065946; PID:g1065949; PIDN:AAA814
C:Genetics:
A:Gene: CESP.D2096.3
A:Introns: 76/1; 100/1; 186/2; 245/3; 384/1; 669/2; 720/2; 737/3; 752/1; 806/2; 862/3; 8

Query Match 21.7%; Score 1416; DB 2; Length 1743;
Best Local Similarity 28.6%; Pred. No. 1e-77;
Matches 385; Conservative 193; Mismatches 433; Indels 334; Gaps 43;

QY 22 SREKQPKSPKRE-----SCGFPRMKNLQWKAQRLH-----LPPVRL 62
DB 397 SNEHNRGAPTTPKRPVFRKERKRTHEETQ-BELRKIWEIEGIGKHSMSQVSGQFPFKSV 455
QY 63 AELHADLKIQRDELAKWKLDGLDDEGEKARLIRNLNVLAKYGLDGK-----KD 115
DB 456 K-----KDENEE-----DGDQDW-IDHLLARNFRLECKKTTDNLHRSPDRS 500
QY 116 ARQVTSNLSLGTQEDGLDRLKELMHKAKTSKFSGEELDKLWREFLHKKVHEYNVL 175
DB 501 EKVIPAKTSKILQVQEQPKMKMGICRIDAPSSAEHLNLSLK-----EWELVEDQKLQ 557
QY 176 LETLSRTEIHNIVSPSLSDIKGVLHSRTELKEKLRN-QGLDLRLRVSHQGYST 234
DB 558 L-SRKAEBELR-----LLH--YTSL--VSSINRKGIOELDGASKKSEIV 596
QY 235 EAEFEPRVIDLWDLAQSANLTDKEAFR-BELKHFEAKIEKHNHYQKLEIAHEKLRH 293
DB 597 EQEFQK-----RLAEIQRSLQVRNEKREQWKS-----MDELTEKGEELLAF 640
QY 294 ABSVGDGER-----VSRGREKHA 311
DB 641 FELTSDGGRNDDYIWTEDDEDDWYPTYPNHPKSLCSQCAPRQORRIAKREREQAK 700
QY 312 LLEGRTKELGYTVKKHQLDLSRISARAEATGAHPRAVP-----TQC----- 358
DB 701 IKVSCLEKEYGIBKRESLRNINCL--KISVVTSPFSLKVPVSDQKTLGCPCTFRCT 757
QY 359 ----- 358
DB 758 FSFFQRPVLEFPFAENCLHGRKVRKQYPVSHLNIIFLANCKQRLKLMFLKKGSTIV 817
QY 359 -----DVPNPRFCAPDKAITQCEARGCCYIPAKOGLQAGQMGQPCFP 405
DB 818 VLLLVFYQNTNSQSDVSKRVDYCPPEGASQACQSGRCIWEAP---SSSPVGTWCY 874
QY 406 PPSYPSYKLENLSSEMGVTAIILTRT-----TPFPFKDILTLRLDVMETENLHF 457
DB 875 P-----TESGTVQSTGTSNLFVLAAKTKNPYGNISPLNVKY-----SSNGATL 918
QY 458 TTKDPANRREYVLETPRVHSRAPSPLYSEF-----SEEPFGVIVHRLDGRVLNNT 511
DB 919 LUTIGNDDRYVPVNPFPKPSSTE---SLKTSGTIGSDVFSFKVTRASTGVALLWDS 975
QY 512 VAPLFPADQLQLSTLSPQYITGLAEHLSPMLSTSWTRITLW---NRDLAPTGA--- 565
DB 976 ICGMQFADKFIQIATVLPKNIYVFGDHIK-KIRNLDORYITWPMFARDIGDPSGALS 1034
QY 566 --NLYGSHFPYALEDGGSAHGVFLNSNAMDVLOPSPALSWRSTGGILDVIFLGPBP 623
DB 1035 TQNLGVHPFYMCIEADGAHGVFLNSNAQEVETGFGPHLLYRTITGGRIDMAFFPGPTP 1094
QY 624 KSVVQYQLDVGVYFPMPPYGLGFLHCRWYGSSTAITROVVENMTFAHPPLDVQNDLDY 683
DB 1095 EQVNVQYLOHIGFFLPFAYWALGYLCRWYGNLDAMKTVISRNQALGIPLDVPYADIDY 1154

QY 684 MDSRDEFTFNKDGFRDPPAMVQELHOGGRRYMMIVDPAISSGGPAGSVRYPYDEGLR---- 739
DB 1155 MNHYEDFT-EGDNWSGFPAYTQOIHAQGLHLIVFDPAVEV-----DYASFORGINADAS 1208
QY 740 -----RGVFTNETGOLPIKGVKWP-STAFPOFTNP--TALAWMEDWVAEF 782
DB 1209 FIEWARDDQVPHNIQDQYPMKNTKIMLGNVWMPDRNTAPPDLPRNNTNNAWGEFAQF 1268
QY 783 HDQVPFDGLWIDMNPSPNFIKRGSD-----GCP-----NNELENPPYVPGVW--- 824
DB 1269 HKTLPFDGWNIDMNPSPNFDGTNTVVEEQLASAKLSCPIITGSDSLDVPFPYPTQAVYQR 1328
QY 825 -GGTLQAATIC---ASSHQFISTHYNLNLYGLTEAIAHSHALVKARGTRPPIVSRSTFA 880
DB 1329 NGEYLFSTKLMGLKTAHR-TRDFYDTKNLYGWSERATYQAIPOVTKRSAVISRSTFP 1387
QY 881 GHGRVAGHWTGDVWSWEQLASVPEILQFNLLGVPLGADVCGFLGNTSELCLVVRWTOL 940
DB 1388 SSGRYGGHMLGDNTRAWGDQLQTSVIGVMEFNMGFPYVVGSDICGPNVGSNEELCLURHQF 1447
QY 941 GAFYFPMENHNSLLSLPQEPYSFSEPAQQAAMRKALTRYALLPHLYTLFHOAHVAGETVA 1000
DB 1448 GAFSEFSDHNSGPNQDQPAWPSVA-NAAKIALTFRYFLFPLYSLYHNAARYGHTVI 1506
QY 1001 RPLFLFEPKDSSTVVDHQLLWGEALLITPVLQAKAEVGYFPPLGTWYDLOTVPIEALG 1060
DB 1507 RPLFPFPKDBETLNISEQFLWGSALMIAPALYQOTSVAHYFSPSDTWYSLQ----- 1558
QY 1061 SLPPPPAAPREPAIHSEGOVW---TLPAPLDTIN-VHLRAGVILIQLOGPGLTTTESROOP 1116
DB 1559 -----PETYQKMGFNDVNAPLSLTLPVVRGCVLPQSPGTTTTASRLSP 1607
QY 1117 MALAVALTKGGEARGELFWDDGESL 1141
DB 1608 FELLITVKTNAASSGDLYYDGGDDL 1632

RESULT 11
S65057
alpha-glucosidase (EC 3.2.1.20) - barley
C:Species: Hordeum vulgare (barley)
C:Date: 20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004
A:Accession: S65057; S65058
R:Tibbot, B.K.; Skadsen, R.W.
Submitted to the EMBL Data Library, March 1995
A:Description: Molecular cloning and characterization of a gibberellin-inducible alpha-g
A:Reference number: S65057
A:Accession: S65057
A:Molecule type: mRNA
A:Residues: 1-877 <TIB>
A:Cross-references: UNIPROT:Q43763; EMBL:U22450; NID:g944900; PIDN:AA02985.1; PID:g94490
R:Tibbot, B.K.; Skadsen, R.W.
Plant Mol. Biol. 30, 229-241, 1996
A:Title: Molecular cloning and characterization of a gibberellin-inducible, putative alpt
A:Reference number: S65058; MUID:96178863; PMID:8616248
A:Accession: S65058
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 85-115, 'S', 117-121, 123-170, 'D', 172-366, 'G', 368-529, 'H', 531-685, 'G', 687-764 <J
A:Cross-references: EMBL:U22450
C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homology
C:Keywords: glycosidase; hydrolase
F:125-769/Domain: sucrose/isomaltase homology <SIM>

Query Match 21.5%; Score 1403.5; DB 2; Length 877;
Best Local Similarity 37.4%; Pred. No. 2e-77;
Matches 331; Conservative 133; Mismatches 327; Indels 95; Gaps 25;

QY 403 CFFPSPSYKLENLSSESMEGYTATLT-----RTTPTFPFKDILTLRLDVMETEN 453
DB 14 CLFAPRLCSSKEGFLAARTVLAIVATMEGALRAEAATGGRSSTGVDQRLAVVYASLETDS 73

QY 454 RLHFTTKDPAARRRYEVPLE-TPR-----VHSRAP-----SPLYSVERS-----EE 492
Db 74 RLVRITDADHPREVPQDIIPRAPGDVLHDAPPASSAPLOGRVLSPPAGSDLVLTVHAS 133
QY 493 PFGVIVHRLDGRVLLNTTVP--LFFADQFLQSLTSLPS--QYITGLAEHLSPMLSLTWS 549
Db 134 PFRFTVSRRTGTDLTDT--APGLVFRDKYLEVTSALPAGRASLYGLGHTKTSFRLRHN 191
QY 550 TRITLWNRDL-APTPGANLYGSHPPFYALBDGSAHGVLNANMDVVLQSPALSWRS 608
Db 192 DSPTLWADIGASVDVNLGSHPPFYMDVRAPGTAHGVLLLSNGMD-VLYGGSYVYKY 250
QY 609 TCGILOVYIFLGPBPSVVOQLDVGVPMPYPWGLGPHLCRWGYSSTAITRQVENVMT 668
Db 251 IGGVLDYFFPAGPNPLAVDQYTLTARPAMPYFWSFGFHQCRGYLVNLSDLERVARYA 310
QY 669 RAHFPLDQVNDLDYMSRDTFENKDF--RDFPAMVQELHOGGRYMMIVDPAISSSG 726
Db 311 KARIPLVEMWTDIDYMDGFKDFTLDRVNTAABLRPFVDRLHNAQKYVLLDPCIRVDP 370
QY 727 PAGESYRYPDGLRRGVFITNETGQPLIGKVPGSTAPPDFTNPTALAWMEDMVAEFDQV 786
Db 371 IDATYGTGVRGMOQDIFL-KRNGTNGVGNVPGDVYFPDFMHPAAAEFWAREISLFRRTI 429
QY 787 PFGLMIDMNEPNSFIRGSDGCPNLEENPPY--VPGVVGGLQAATI-CASSHQFLST 843
Db 430 PVDGLMIDMNEISNFNPE-----PMNALDDPPYRINNDGTGRPINNKTRVPLAVHYGGVT 485
QY 844 HYNLNLGYLGEAASHRALVKARGTRPFVISRSTFAGHRYAGHTGDMVSSWEOLASS 903
Db 486 EYEBHNLFGLEARATGRVLRDTRRRPFVLSRSTFVSGRYTAYWTGDNAAATWGDLRYS 545
QY 904 VPBILFNLGLVPLGADVCGFLGNTSEELCVRWTLQLGAPYPMRNHNSLLSLPQPSYF 963
Db 546 INTMLSFGLGMPWIGADICGPNGTTEELCGRWIQLGAPYPSRHSALFTVRRELYLW 605
QY 964 SEPAQAMKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTDHQLLWG 1023
Db 606 PSVAASG-RKALGRYQLLFYFTLYMEAHMTGAPARPLFFSYPHDVATYGVDRQFLLG 664
QY 1024 EALLITPVLOAGKAEVTGVPPLGTWVDLQTVPIEALGSLPPPPAAREPAIHSEGOVTL 1083
Db 665 RGVLSFVLEPGPTTVDVAPFAGRWLRLYDYSL-----AVATR-----TGRHRL 709
QY 1084 PAPLDTINVHLRAGYIPIQPGGLTTESTRQOPMALAVALKGEARGELFMDGDSLEV 1143
Db 710 PAPADTVNHLTGTLPLQOQSALTTSRARTAFHLLVALAEDGTASGYLFLDDGDSPEY 769
QY 1144 LERGATQVIFLARNNTVNE--LVRTSE-----GAGLOQKVTLG--VATAPOQ 1191
Db 770 GRRSDSMVRF---NYKIPNKGAIKVKSEVHNHYAQSRTLVISKVLMGHRSPAAPKK 826
QY 1192 ----VLSNGVPSNFTYSPDTKVDI-----CVSLLMGEQF 1223
Db 827 LTVHNSAEVEASS---SAGRIQNAAGLGGVVAHIGLSLVGGEF 869

RESULT 12
T38598
Probable family 31 glycosyl hydrolase (alpha glucosidase) precursor - fission yeast (Sch
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence revision 20-Oct-2000 #text_change 09-Jul-2004
C:Accession: T38598; S62559; T38911
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z21801
A:Accession: T38598
A:Molecule type: DNA
A:Residues: 1-385 <PEA>
A:Cross-references: UNIPROT:Q09901; EMBL:Z67961; PIDN:CAA91887.1; GSPDB:GN00066; SPDB:SH
A:Experimental source: strain 972h-; cosmid c30D11
R:Pearson, D.; Churcher, C.M.
submitted to the EMBL Data Library, November 1995

A:Reference number: S62559
A:Accession: S62559
A:Molecule type: DNA
A:Residues: 1-384 <PE2>
A:Cross-references: EMBL:Z67961
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38911
A:Molecule type: DNA
A:Residues: 352-993 <PE3>
A:Cross-references: EMBL:Z69728; PIDN:CAA93572.1; GSPDB:GN00066; SPDB:SPAC56F8.01
A:Experimental source: strain 972h-; cosmid c56F8
C:Genetics:
A:Gene: SPDB:SPAC30D11.01c; SPDB:SPAC56F8.01
A:Map position: 1L
C:Superfamily: Schwanmioncyes glucan 1,4-alpha-glucosidase GAM1; sucrase/isomaltase homo
F:165-384/Domain: sucrase/isomaltase homology (fragment) <SIM>

Query Match 21.2%; Score 1380; DB 2; Length 993;
Best Local Similarity 34.7%; Pred. No. 6.6e-76;
Matches 328; Conservative 142; Mismatches 317; Indels 158; Gaps 28;

QY 408 SYPSYKLENLSSENGYATLRTTPTTFP--KDIILTLDVMMETENRLHFTIKD----- 461
Db 81 SCFGYQARNISYSYGVLAILELAGDACYAGTYDYPYLLNVNSYDEERVHISISDLNQT 140
QY 462 ---PANRR--YVPLETPRVHSRAPSPLYSVFSEBPPGVIVHRLDGRVLLNTTVPAPLF 516
Db 141 QQLSNRRDWDAPL-FYRSSNFSGNLQYNFNTDPPFWITRIADDQVLPTRGNPLI 199
QY 517 FADQFLQLSTLPSQY-ITGLAEHLSPMLSTSWTRITLWNRDLAFTPGANLYGSHPPYL 575
Db 200 FEDQVIELTNNVEDYVNYVLSGQSQSPRLGNLTK-TFWATGYSQSPSEANMYGSHPPY 258
QY 576 A---LEDG-----GSAHGVFLNANMDVQLQSPALSWRSTGGTLDVYIFLGP--EPK 624
Db 259 EORYIPIGTNTYTASGHVLMSSNGMEVLLR-STYIKYRMIGGIIDLVFVSGSTVSPK 317
QY 625 SVVQVQYLDVVGYPMPYMGFLGFLCRWGYSTAITRQVVENNTRAHFPLDQVNDLDY 684
Db 318 YTIQYVQSIGTPTWQPYWSLGFQMSRWGYKTLSDLINRSLYNASNITEGFMNDIDYM 377
QY 685 DSRRTDFTNKDGF-----RDFPAMVQELHOGGRYMMIVDPAISSSGPAGS-----YRPYD 735
Db 378 SEFRFTVNSTAFPPNQTLDFRSLDSEHQ---HYVPVLDPAIYAANPNKSAADRTYPPY 434
QY 736 EGLRGVFTTNETGQPLIGKVPGSTAPDFTNPTALAWMEDMV-----AEFHDOVP 787
Db 435 SGFEDNIFIKPNPGSAYVGMWPGFVYVPDFTNPAPVLYQWKGILNLSTAFGSNYSYDLP 494
QY 788 FDGLWIDMNEPNSFIRGSDGCPNNELE---NPPY-VPGVVGGLT----- 828
Db 495 FSGLCIDMNEPTSFICGS---CGSLLKLNLPVHPFSLPGDQVNDKVKYSPEDFNATNTE 551
QY 829 -----QAATICASSHQFLS-----TH----- 844
Db 552 YKSVSRASQSKATATSEKSHETPSSESLINGKPEFSINYPYALDITDTHDLAQFGV 611
QY 845 -----YNLNLGYLGEAASHRALVKAR-GTRPFFVISRSTFAGHRYAGHTG 891
Db 612 SPNATMHGNTLYNLFNTYGYSEKISFSEALNSIQNIRPPLLSRSTFVSGRYAAHVLG 671
QY 892 DYSSWEOLASSVPEILOENLGLVPLGADVCGFLGNTSEELCVRWTLQLGAPYPMRNHN 951
Db 672 DNKSQSDMVSSISILFILLGIPMGADVCGYNGNTDEELCARWALGAFLPFRNHN 731
QY 952 SLLSLPQEPYSPSEPAQAMKALTLRYALLPHLYTLFHOAHVAGETVARPLFLFPKDS 1011
Db 732 SLGSIQEPPEFRWASVA-EASRSAIBIRYSLLPYWTLMTHTASVDGTPMVRPLFFEPKQI 790
QY 1012 STWTVDHQLLWGEALLITPVLOAGKAEVTGYFPLGT---WYDLQTVPIEALGSLPPPPAA 1068

791 SLASVDKQFMIGTALLISPALENTYIIQIGIIGDNDTIWYDWNHVSU----- 838

1069 PREPAIHSEQWVTLTPAPLDTINVHLRAGYIIPLQPGPLTTTTSRQOPMALAVALTGGGE 1128

839 -----INHVDENITMSAPLCYVNIARVGGNIIPLOQPGYTTYSRNPYSLLIAMDNNGF 894

1129 ARGELFWDDGESLEVLERGAYTOVIFLARNTI--VNELVRVTSEGAGLQKVTVLGVA 1186

895 ASGSLYIDDDIGSMQT--NSSLSVKLNSNNTITCVVSGTMWSSP-----SLANITITGLS 947

1187 TAPOQVLSNGVPVSNFTYSPD-----TKVLDICVSLLMGEQFLYSW 1227

948 NPENTILFNQQLSDYQYSQTLSTLNLDDLTVDGAFSKNWTVTW 992

RESULT 13

T50267

probable family 31 glucosidase [imported] - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004

C:Accession: T50267

R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, December 1999

A:Reference number: Z25031

A:Accession: T50267

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-995 <HUN>

A:Cross-references: UNIPROT:Q9URX4; EMBL:AL133522; PIDN:CAB63549.1; GSPDB:GN000666; SPDB:SPAC1039.11c; SPDB:SPAC922.02c

A:Gene: SPAC1039.11c; SPDB:SPAC922.02c

A:Map position: 1

C:Superfamily: Schwanniomycetes glucan 1,4-alpha-glucosidase GAML1; sucrase/isomaltase homologue

Query Match 20.9%; Score 1361.5; DB 2; Length 995;

Best Local Similarity 35.3%; Pred. No. 8.9e-75;

Matches 328; Conservative 129; Mismatches 305; Indels 167; Gaps 28;

QY 412 YKLENLSSSEMGYATIL-TRTTPTF-PPKDILRLDVMMETENRHLFTIKDPANRREYV 469

DB 79 YNAIQVSEKGVAILQLQNGDPCYAGTDYPPFLALNVSPFSDIRLHVSIQDLYGAQOF 138

QY 470 PLET-----PRVHSRAP--SPLYSVFSEBEPFGVIVHRQLDGRVLLNTTVAPLFFAQFL 522

DB 139 SKRTDWDAPLYHFQFGDRTYNFSNQPFEFWTRVSDGVLFDTRGHKLIFEDQYI 198

QY 523 QLSTSLPSQY-ITGLAEHLSPMLSTSWTRITLWNRDLAPT-GANLYGSHPPYL----- 575

DB 199 ELATTNVDDYNYVGLAETVHGLRLGNLIR-TPWANG-NPTPLDRNAYGTHPPYLSHRYT 256

QY 576 ----ALEDG-----GSAHGCVLLNSAMDVVLOPSALSWSRSTGGILDVYIFLG--PEPK 624

DB 257 PSENLSNDGQPSYTSSTGHVLMITANGMEVLLRPN-YLQYRIIGGIVDLYIYVGGTKNPK 315

QY 625 SVVQQVLDVVGYPFMPYVWGLGPHLCRWGYSSTAITRQVVENMTRAHFFLDVQVNDLDYM 684

DB 316 DTVSQFVQSGVTAMQOHWTFGPHICRWGYNVFDLVEVKNFKNFEIPVDVFWSDIDYM 375

QY 685 DSRDDFT-----FNKQGRFDPFAMVQELHGQGRYRMYIVDPAISSGSPAGS-----YRPYD 735

DB 376 YEYRDFTVESNAPFKDKWMEF---FNSLQSQNHVYPIIDAALYAANPINRSDDVYVPY 432

QY 736 EGLRGVFTNETGQPLIGKWVGSTAFDPFTNPTALAWMEDVMAEFD-----QVP 787

DB 433 EGVRRDIFLRNPORSLYVGNWVFGFTTFDDFTNPETNTYVTECLMNSAAGYNSFPFLP 492

QY 788 FDLGLWIDMNPSPNFRIGSGDCGNNELENPPYVGVV-----GGTL 828

DB 493 YSLGLWIDMEPTSCFICGS---CGTDKLDQNPVHPALIEGEPNNMVMYPEGEHTNASE 549

QY 829 QAATICASSHQFLUST----- 843

This Page Blank (uspto)

MEDLINE=93168114; PubMed=8435067;

RA Hermans M.M.P., Wisselaar H.A., Kroos M.A., Oostra B.A.,
RA Reuser A.J.J.;
RT "Human lysosomal alpha-glucosidase: functional characterization of the
RT glycosylation sites";
RL Biochem. J. 289:681-686 (1993).
RN [17]
RN CARBOHYDRATE-LINKAGE SITE ASN-470.
RX MEDLINE=22660472; PubMed=12754519; DOI=10.1038/nbt827;
RA Zhang H., Li X.-J., Martin D.B., Abersold R.;
RA "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry";
RL Nat. Biotechnol. 21:660-666 (2003).
RN [18]
RN REVIEW ON VARIANTS.
RX MEDLINE=95327152; PubMed=7603530;
RA Reuser A.J.J., Kroos M.A., Hermans M.M.P., Bijvoet A.G.A.;
RA Verbeet M.P., van Diggelen O.P., Kleijer W.J., van der Ploeg A.T.;
RT "Glycogenosis type II (acid maltase deficiency).";
RL Muscle Nerve 3:S61-S69 (1995).
RN [9]
RN VARIANT ASN-91.
RX MEDLINE=90365036; PubMed=2203258;
RA Martinik F., Bodkin M., Tzall S., Hirschhorn R.;
RT "Identification of the base-pair substitution responsible for a human
RT acid alpha glucosidase allele with lower 'affinity' for glycogen (GAA
RT 2) and transient gene expression in deficient cells.";
RL Am. J. Hum. Genet. 47:440-445 (1990).
RN [10]
RN VARIANT GSD-II THR-318.
RX MEDLINE=91353580; PubMed=1652892;
RA Zhong N., Martinik F., Tzall S., Hirschhorn R.;
RT "Identification of a missense mutation in one allele of a patient with
RT Pompe disease, and use of endonuclease digestion of PCR-amplified RNA
RT to demonstrate lack of mRNA expression from the second allele.";
RL Am. J. Hum. Genet. 49:635-645 (1991).
RN [11]
RN VARIANT GSD-II LYS-521.
RX MEDLINE=91379015; PubMed=1898413;
RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A., Oostra B.A.,
RA Reuser A.J.J.;
RT "Identification of a point mutation in the human lysosomal alpha-
RT glucosidase gene causing infantile glycogenosis type II.";
RL Biochem. Biophys. Res. Commun. 179:919-926 (1991).
RN [12]
RN VARIANTS GSD-II ARG-643 AND TRP-725.
RX MEDLINE=94004908; PubMed=8401535;
RA Hermans M.M.P., Kroos M.A., de Graaff E., Oostra B.A., Reuser A.J.J.;
RT "Two mutations affecting the transport and maturation of lysosomal
RT alpha-glucosidase in an adult case of glycogen storage disease type
RT II.";
RL Hum. Mutat. 2:268-273 (1993).
RN [13]
RN VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.
RX MEDLINE=93168115; PubMed=8094613;
RA Hermans M.M.P., de Graaff E., Kroos M.A., Wisselaar H.A.,
RA Willemssen R., Oostra B.A., Reuser A.J.J.;
RT "The conservative substitution Asp-645-->Glu in lysosomal alpha-
RT glucosidase affects transport and phosphorylation of the enzyme in an
RT adult patient with glycogen-storage disease type II.";
RL Biochem. J. 289:687-693 (1993).
RN [14]
RN VARIANT GSD-II GLU-645, AND VARIANTS ILE-816 AND ILE-927.
RX MEDLINE=92096118; PubMed=1684505;
RA Martinik F., Mehler M., Bodkin M., Tzall S., Hirschhorn K., Zhong N.,
RA Hirschhorn R.;
RT "Identification of a missense mutation in an adult-onset patient with
RT glycogenosis type II expressing only one allele.";
RL DNA Cell Biol. 10:681-687 (1991).
RN [15]
RN VARIANTS ILE-816 AND ILE-927.
RX MEDLINE=93252406; PubMed=8486380;
RA Hermans M.M.P., Svetkey L.P., Oostra B.A., Chen Y.T., Reuser A.J.J.;
RT "The loss of a polymorphic glycosylation site caused by Thr-927-->Ile
is linked to a second polymorphic Val-816-->Ile substitution in
lysosomal alpha-glucosidase of American blacks.";
RL Genomics 16:300-301 (1993).
RN [16]
RN VARIANT GSD-II VAL-519.
RX MEDLINE=95170739; PubMed=7866409;
RA Huie M.L., Hirschhorn R., Chen A.S., Martinik F., Zhong N.;
RT "Mutation at the catalytic site (M519V) in glycogen storage disease
RT type II (Pompe disease).";
RL Hum. Mutat. 4:291-293 (1994).
RN [17]
RN VARIANT GSD-II TRP-647.
RX MEDLINE=95072571; PubMed=7981676;
RA Huie M.L., Chen A.S., Brooks S.S., Grix A., Hirschhorn R.;
RT "A de novo 13 nt deletion, a newly identified C647W missense mutation
RT and a deletion of exon 18 in infantile onset glycogen storage disease
RT type II (GSDII).";
RL Hum. Mol. Genet. 3:1081-1087 (1994).
RN [18]
RN VARIANT GSD-II LEU-545.
RX MEDLINE=95187163; PubMed=7881422;
RA Hermans M.M.P., de Graaff E., Kroos M.A., Mohkamsing S., Eussen B.J.,
RA Joosse M., Willemssen R., Kleijer W.J., Oostra B.A., Reuser A.J.J.;
RT "The effect of a single base pair deletion (delta 1525) and a C1634T
RT missense mutation (Pro545Leu) on the expression of lysosomal alpha-
RT glucosidase in patients with glycogen storage disease type II.";
RL Hum. Mol. Genet. 3:2213-2218 (1994).
RN [19]
RN VARIANTS GSD-II ARG-299; LYS-903 DEL AND VARIANTS HIS-199; ARG-223 AND
RX VAL-780.
RX MEDLINE=95233437; PubMed=7717400;
RA Boerkoel C.F., Exelbert R., Nicastri C., Nichols R.C., Miller F.W.,
RA Plotz P.H., Raben N.;
RT "Leaky splicing mutation in the acid maltase gene is associated with
RT delayed onset of glycogenosis type II.";
RL Am. J. Hum. Genet. 56:887-897 (1995).
RN [20]
RN SEQUENCE OF 631-680 FROM N.A., AND VARIANT GSD-II HIS-645.
RX MEDLINE=95209708; PubMed=7695647;
RA Lin C.-Y., Shieh J.-J.;
RT "Identification of a de novo point mutation resulting in infantile
RT form of Pompe's disease.";
RL Biochem. Biophys. Res. Commun. 208:886-893 (1995).
RN [21]
RN VARIANT LYS-689.
RX PubMed=8912788;
RA Huie M.L., Menaker M., McAlpine P.J., Hirschhorn R.;
RT "Identification of an E689K substitution as the molecular basis of the
RT human acid alpha-glucosidase type 4 allozyme (GAA*4).";
RL Ann. Hum. Genet. 60:365-368 (1996).
RN [22]
RN VARIANT GSD-II VAL-529.
RX MEDLINE=96431168; PubMed=8834250;
RA Tsunoda H., Ohshima T., Tohyama J., Sasaki M., Sakuragawa N.,
RA Martinik F.;
RT "Acid alpha-glucosidase deficiency: identification and expression of a
RT missense mutation (S529V) in a Japanese adult phenotype.";
RL Hum. Genet. 97:496-499 (1996).
RN [23]
RN VARIANTS GSD-II ASN-645; TRP-647; SER-648; GLN-672 AND TRP-672.
RX MEDLINE=98205825; PubMed=9535769; DOI=10.1006/bbrcl.1998.8255;
RA Huie M.L., Tsujino S., Brooks S.S., Engel A., Elias E., Bonthron D.T.,
RA Bessley C., Shanske S., Dimauro S., Goto Y.I., Hirschhorn R.;
RT "Glycogen storage disease type II: identification of four novel
RT missense mutations (D645N, G648S, R672W, R672Q) and two
RT insertions/deletions in the acid alpha-glucosidase locus of patients
RT of differing phenotype.";
RL Biochem. Biophys. Res. Commun. 244:921-927 (1998).
RN [24]
RN VARIANT GSD-II ARG-309.
RX PubMed=9660056;
RA Kroos M.A., van Leenen D., Verbiest J., Reuser A.J.J., Hermans M.M.P.;
RT "Glycogen storage disease type II: identification of a dinucleotide

RT deletion and a common missense mutation in the lysosomal alpha-glucosidase gene.";

```
Query Match      72.7%; Score 4744.5; DB 1; Length 952;
Best Local Similarity 97.2%; Pred. No. 1.7e-264;
Matches 890; Conservative 3; Mismatches 18; Indels 5; Gaps 1;

QY 313 LEGRTKELGYTVKKHLQDLISGRISARABAEAGHPRPRAVPTQCDVPNSRFDCAADK 372
DB 42 LSGSSPVLEETHPAHQGASRPGPR-----DAQAHGPRPRAVPTQCDVPNSRFDCAADK 96

QY 373 AITQCEARGCCYIPAKQLOGAQMGQPCWCPFPSPYSYKLENLSSEMGYTATLTRTT 432
DB 97 AITQCEARGCCYIPAKQLOGAQMGQPCWCPFPSPYSYKLENLSSEMGYTATLTRTT 156

QY 433 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 492
DB 157 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 216

QY 493 PFGVIVRHQLDGRVLLNTTVAFLFFADQFLQSLTSPSQYITGLAEHLSPMLSTSWTRI 552
DB 217 PFGVIVRHQLDGRVLLNTTVAFLFFADQFLQSLTSPSQYITGLAEHLSPMLSTSWTRI 552

QY 553 TLNWRDLATPGANLYGSHPPYLALEDGGSAGHGVFLNNSNAMDVVLQSPALSWRSTGGI 612
DB 277 TLNWRDLATPGANLYGSHPPYLALEDGGSAGHGVFLNNSNAMDVVLQSPALSWRSTGGI 336

QY 613 LDVYIFLGPPEKSVQOYLVDVGYPPMPYVWGLGFLHCRWYGSSTAIITQVVENMTTRAHF 672
DB 337 LDVYIFLGPPEKSVQOYLVDVGYPPMPYVWGLGFLHCRWYGSSTAIITQVVENMTTRAHF 396

QY 673 PLDVQWNLDDYMSRRDFTFNKGDFRDFPAMVQELHQGGRYNNMVDPAISSSGPAGSYR 732
DB 397 PLDVQWNLDDYMSRRDFTFNKGDFRDFPAMVQELHQGGRYNNMVDPAISSSGPAGSYR 456

QY 733 PYDEGLRGVFTNETGQPLIGKVGSTAFPPDTNPTALAWWEDVVAEFHDQVPDGLW 792
DB 457 PYDEGLRGVFTNETGQPLIGKVGSTAFPPDTNPTALAWWEDVVAEFHDQVPDGLW 516

QY 793 IDWNEPSNFTIRGSEDCGPNNELENPPVPGVVGTTLOAATTCASSHQFLSTHYNLNLVY 852
DB 517 IDWNEPSNFTIRGSEDCGPNNELENPPVPGVVGTTLOAATTCASSHQFLSTHYNLNLVY 576

QY 853 LTAIAASHRALVKARGTRPVIISRTFAGHGRYAGHWTGDMVSWBQLASSVPEILQFNL 912
DB 577 LTAIAASHRALVKARGTRPVIISRTFAGHGRYAGHWTGDMVSWBQLASSVPEILQFNL 636

QY 913 LGVPLVGVADVCGFLGNTSELCVRWTQLGAFYFPFMRNHNLSLDPQEPYFSFSPAQAAMR 972
DB 637 LGVPLVGVADVCGFLGNTSELCVRWTQLGAFYFPFMRNHNLSLDPQEPYFSFSPAQAAMR 696

QY 973 KALTLYALLPHLYTLFQAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPVL 1032
DB 697 KALTLYALLPHLYTLFQAHVAGETVARPLFLEFPKDSSTWTDHQLLWGEALLITPVL 756

QY 1033 QAGKAETVGYFPLGTWYDLQTVIEALGSLPPPPAAPREPAIHSEGOVTLPLADTINV 1092
DB 757 QAGKAETVGYFPLGTWYDLQTVIEALGSLPPPPAAPREPAIHSEGOVTLPLADTINV 816

QY 1093 HLRAGYIIPLOQPLGTTTSSRQOPMALAVATKGGEGARGELFWDGSELEVLGRGAYTV 1152
DB 817 HLRAGYIIPLOQPLGTTTSSRQOPMALAVATKGGEGARGELFWDGSELEVLGRGAYTV 876

QY 1153 IFLARNNTIINELVRVTSEAGLQKQVTVLGATAPQOVLNGVPVSNFTYSPDTKVLVD 1212
DB 877 IFLARNNTIINELVRVTSEAGLQKQVTVLGATAPQOVLNGVPVSNFTYSPDTKVLVD 936

QY 1213 ICVSLLMGEQFLVSWC 1228
DB 937 ICVSLLMGEQFLVSWC 952
```

RESULT 2

```
Q81WE7 PRELIMINARY; PRT; 952 AA.
AC Q81WE7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE GAA protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
SEQUENCE FROM N.A.
RP TISSUE=DuoDenum;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
RP TISSUE=DuoDenum;
RX Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RA Strausberg R.;
RL EMBL; BC040431; AAH40431.1; -
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
SQ SEQUENCE 952 AA; 105323 MW; 6E2717BF7201F469 CRC64;

Query Match      72.5%; Score 4730.5; DB 2; Length 952;
Best Local Similarity 96.8%; Pred. No. 1.1e-263;
Matches 887; Conservative 4; Mismatches 20; Indels 5; Gaps 1;

QY 313 LEGRTKELGYTVKKHLQDLISGRISARABAEAGHPRPRAVPTQCDVPNSRFDCAADK 372
DB 42 LSGSSPVLEETHPAHQGASRPGPR-----DAQAHGPRPRAVPTQCDVPNSRFDCAADK 96

QY 373 AITQCEARGCCYIPAKQLOGAQMGQPCWCPFPSPYSYKLENLSSEMGYTATLTRTT 432
DB 97 AITQCEARGCCYIPAKQLOGAQMGQPCWCPFPSPYSYKLENLSSEMGYTATLTRTT 156

QY 433 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 492
DB 157 PTFFPKDILTLRLDVMMETENRLHFTIKDPANRRYEVPLETPRVHGRAPSPLYSVFSEE 216

QY 493 PFGVIVRHQLDGRVLLNTTVAFLFFADQFLQSLTSPSQYITGLAEHLSPMLSTSWTRI 552
DB 217 PFGVIVRHQLDGRVLLNTTVAFLFFADQFLQSLTSPSQYITGLAEHLSPMLSTSWTRI 552
```

```
QY 553 TLNWRDLAPTPGANLYGSHHPFYALBEGGSAHGVFLNNSNAMDVILQSPALSWRSTGGI 612
Db TLNWRDLAPTPGANLYGSHHPFYALBEGGSAHGVFLNNSNAMDVILQSPALSWRSTGGI 612
QY 613 LDVYIFLGPEPKSVQOYLDDVGVCPWMPYVWGLGFLCRWGYSSSTAITRQVVENMTTRAHP 672
Db LDVYIFLGPEPKSVQOYLDDVGVCPWMPYVWGLGFLCRWGYSSSTAITRQVVENMTTRAHP 672
QY 673 PLDVQWNLDDYMDSRDFTFNKDGFRDFPAMVQELHGGRRYMMIVDPAISSGGPAGSYR 732
Db PLDVQWNLDDYMDSRDFTFNKDGFRDFPAMVQELHGGRRYMMIVDPAISSGGPAGSYR 732
QY 733 PYDEGLRGGVFIINETGQPLIGKVPKSTAFPDFTNPTALAMWEDMVAFHQVDPDGLW 792
Db PYDEGLRGGVFIINETGQPLIGKVPKSTAFPDFTNPTALAMWEDMVAFHQVDPDGLW 792
QY 793 IDWNEPSNFIIRGSEDCPNNELENPPYVGVVGGTLOAATICASSHQFLSTHYNLNLGY 852
Db IDWNEPSNFIIRGSEDCPNNELENPPYVGVVGGTLOAATICASSHQFLSTHYNLNLGY 852
QY 853 LTBIAIASHRALVKARGTRPFVSRSTFAGHGRVAGHWTDGVWSSWEQLASSVPEILQFNL 912
Db LTBIAIASHRALVKARGTRPFVSRSTFAGHGRVAGHWTDGVWSSWEQLASSVPEILQFNL 912
QY 913 LGVPLVGVADVCGFLGNTSBEELCVRWTLQGFYFPMRNHNSLSLPOEPYSFSEPAQAMR 972
Db LGVPLVGVADVCGFLGNTSBEELCVRWTLQGFYFPMRNHNSLSLPOEPYSFSEPAQAMR 972
QY 973 KALTIRYALLPHLYTLFQAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 1032
Db KALTIRYALLPHLYTLFQAHVAGETVARPLFLEFPKDSSTWVDHQLLWGEALLITPVL 1032
QY 1033 QAGKAEVTVGFLGTWYDQVPIEALGSLPPPPAAPREPAHSEGOVWTLPAPLDTINV 1092
Db QAGKAEVTVGFLGTWYDQVPIEALGSLPPPPAAPREPAHSEGOVWTLPAPLDTINV 1092
QY 1093 HLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTV 1152
Db HLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGEARGELFWDDGESLEVLERGAYTV 1152
QY 1153 IFLARNNTIVNELVRVTSAGLOLOKVTVLGVATAPQOVLNGVPSNFTYSPDTKVL 1212
Db IFLARNNTIVNELVRVTSAGLOLOKVTVLGVATAPQOVLNGVPSNFTYSPDTKVL 1212
QY 1213 ICVSLLMGEQFLVSWC 1228
Db ICVSLLMGEQFLVSWC 952
QY 937 ICVSLLMGEQFLVSWC 952
Db ICVSLLMGEQFLVSWC 952
RESULT 3
Q9MTM4
ID Q9MYM4 PRELIMINARY; PRT; 937 AA.
AC Q9MYM4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Acidic alpha-glucosidase.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20188772; PubMed=10723725;
RA Dennis J.A., Moran C., Healy P.J.;
RT "The bovine alpha-glucosidase gene: coding region, genomic structure,
RL Mamm. Genome 11:206-212(2000)."
DR EMBL; AF171666; AAF81637.1; -
DR EMBL; AF171665; AAF81636.1; -
DR HSSP; P01359; 2SP.
DR Go; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
```

```
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil_1.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil_1.
DR SMART; SM00018; FD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; UNKNOWN 1.
SQ SEQUENCE 937 AA; 104756 MW; B7E9A2D66BAF5026 CRC64;

Query Match 62.2%; Score 4057.5; DB 2; Length 937;
Best Local Similarity 85.0%; Pred. No. 6.2e-225;
Matches 747; Conservative 60; Mismatches 69; Indels 3; Gaps 2;

QY 349 GRPRAVPTQCVPPNSRFDCAKAITOQCEARGCCVIPAQGLQGAGMGPWCFFPFS 408
Db GRPRAVPTQCVPPNSRFDCAKAITOQCEARGCCVIPAQGLQGAGMGPWCFFPFS 408
QY 409 YPSYKLENLSSEMGTATLTRTTPFPKQILRLDLVMMETENRLHFTIKDPANRRYE 468
Db YPSYKLENLSSEMGTATLTRTTPFPKQILRLDLVMMETENRLHFTIKDPANRRYE 468
QY 469 VPLETPRVHSRAPSPLYSVSEFSEEPGVIVRQLDGRVLLNTTVAPLFPADQFLQSTSL 528
Db VPLETPRVHSRAPSPLYSVSEFSEEPGVIVRQLDGRVLLNTTVAPLFPADQFLQSTSL 528
QY 529 PSQYITGLAEHLSPMLSTSWTRITLWNRDLAPTGANLYGSHHPFYALBEGGSAHGVFL 588
Db PSQYITGLAEHLSPMLSTSWTRITLWNRDLAPTGANLYGSHHPFYALBEGGSAHGVFL 588
QY 589 LNSNAMDVVLQSPALSWRSTGGILDVYIFLGPEPKSVVQYLDVVGYPFMPYMGGLGFH 648
Db LNSNAMDVVLQSPALSWRSTGGILDVYIFLGPEPKSVVQYLDVVGYPFMPYMGGLGFH 648
QY 649 LCRWGYSSSTAITRQVVENMTTRAHPFLDVQWNLDDYMDSRDFTFNKDGFRDPAMVQELH 708
Db LCRWGYSSSTAITRQVVENMTTRAHPFLDVQWNLDDYMDSRDFTFNKDGFRDPAMVQELH 708
QY 709 QGGRYVMIIVDPAISSGPAGSYRYPDEGLRGGVFIINETGQPLIGKVPKSTAFPDFTN 768
Db QGGRYVMIIVDPAISSGPAGSYRYPDEGLRGGVFIINETGQPLIGKVPKSTAFPDFTN 768
QY 769 PTALAWEDMVAFHQVDPFDFGLWIDMNEPSNFIIRGSEDCPNNELENPPYVGVVGGTL 828
Db PTALAWEDMVAFHQVDPFDFGLWIDMNEPSNFIIRGSEDCPNNELENPPYVGVVGGTL 828
QY 829 QAAATICASSHQFLSTHYNLNLGYLGLTEALASHRALVKARGMRPFVSRSTFAGHGRYSGH 888
Db QAAATICASSHQFLSTHYNLNLGYLGLTEALASHRALVKARGMRPFVSRSTFAGHGRYSGH 888
QY 889 WTGDDVSSWEQLASSVPEILQFNLGVLGVADVCGFLGNTSEELCVWTLQGAFFPMMR 948
Db WTGDDVSSWEQLASSVPEILQFNLGVLGVADVCGFLGNTSEELCVWTLQGAFFPMMR 948
QY 949 NHNSLSLQSPYSPSEPAQAMRKALTRVALLPHLYTLFQAHVAGETVARPLFLEPP 1008
Db NHNSLSLQSPYSPSEPAQAMRKALTRVALLPHLYTLFQAHVAGETVARPLFLEPP 1008
QY 1009 KDSSTWTVVDHQLLWGEALLITPVLQAGKAEVTVGYPPLGTWYDQVPIEALGSLPPPPAA 1068
Db KDSSTWTVVDHQLLWGEALLITPVLQAGKAEVTVGYPPLGTWYDQVPIEALGSLPPPPAA 1068
QY 1069 PREPAHSEGOVWTLPAPLDTINVHLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGE 1128
Db PREPAHSEGOVWTLPAPLDTINVHLRAGYIIPLOQGLTTTSSRQOPMALAVALTKGGE 1128
QY 1129 ARGELFWDDGESLEVLERGAYTVQVIFLARNNTIVNELVRVTSAGLOLOKVTVLGVATAP 1188
Db ARGELFWDDGESLEVLERGAYTVQVIFLARNNTIVNELVRVTSAGLOLOKVTVLGVATAP 1188
QY 838 AQGELFWDDGESLEVLERGAYTVQVIFLARNNTIVNELVRVTSAGLOLOKVTVLGVATAP 897
Db AQGELFWDDGESLEVLERGAYTVQVIFLARNNTIVNELVRVTSAGLOLOKVTVLGVATAP 897
QY 1189 PQOVLNGVPSNFTYSPDTKVLDCVSLLMGEQFLVSW 1227
Db PQOVLNGVPSNFTYSPDTKVLDCVSLLMGEQFLVSW 1227
```

```

Db      898  PQQVLCNSVPSNFTSPDTETLAIPVSLTMGSEQFVSW 936
RESULT 4
Q6P7A9
ID      Q6P7A9          PRELIMINARY;          PRT;      953 AA.
AC      05-JUL-2004 (TrEMBLrel. 27, Created)
DT      05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT      05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE      Glucosidase, alpha, acid (Pompe disease, glycogen storage disease type
DE      II).
GN      Name=Gaa;
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
SEQUENCE FROM N.A.
RC      TISSUE=Prostate;
RX      MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA      Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA      Altschul S.F., Zerbini B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA      Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA      Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA      Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA      Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA      Fahey J., Helton E., Kettaman M., Madan A.C., Rodriguez S., Sanchez A.,
RA      Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA      Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA      Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA      Krzywinski M.I., Skaleka U., Smailus D.E., Schnerk A., Schein J.E.,
RA      Jones S.J., Marra M.A.;
RT      "Generation and initial analysis of more than 15,000 full-length human
RT      and mouse cDNA sequences";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
RC      TISSUE=Prostate;
RX      Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RA      Strausberg R.;
DR      EMBL; BC061753; AAH61753.1; -
DR      GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR      GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR      InterPro; IPR000322; Glyco hydro_31.
DR      InterPro; IPR000519; P_trefolil.
DR      Pfam; PF01055; Glyco_hydro_31; 1.
DR      Pfam; PF00088; Trefolil; 1.
DR      SMART; SM00018; PD; 1.
DR      PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR      PROSITE; PS00127; GLYCOSYL_HYDROL_F31_2; 1.
DR      PROSITE; PS00025; P_TREFOIL; 1.
SQ      SEQUENCE 953 AA; 106206 MW; D99723779AD41E05 CRC64;

Query Match      62.1%; Score 4047; DB 2; Length 953;
Best Local Similarity      83.0%; Pred. No. 2.6e-224;
Matches 752; Conservative 54; Mismatches 94; Indels 6; Gaps 2;

QY      323  TVKKHQLDLSGRISRAAEATCAHGRPRNAVPTQCDVPNSRFDCAKATQEOCEAR 382
DB      52  TYRPHHQE-----SYRPAHLIQHAEARVPTQCDVTNPSRFDCAKATQEOCEAR 106
QY      383  GCYIPAKOGLQAGQWQPCFPFPPSPYKLNLSSEMGYATATLTRTPFPFKDILT 442
DB      107  GCCWVAGQVLGVGVQWQPCFPFPPSPYKLNLSSEMGYATATLTRTPFPFKDILT 166
QY      443  LRLDVNMETENRLHFTIKDPANRRYVPLETPRVHSRAPSPLYSVFSEEPFGVIVRRQL 502

```

```

RT lysosomal alpha-glucosidase.";
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Heart, and Thymus;
RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nakai I., Otsu T., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmer S., Gustinchin S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierzki R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Malcais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino K., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Breast tumor;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Luetjens J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Essential for the degradation of glycogen to glucose in
CC lysosomes.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing 1,4-
CC linked D-glucose residues with release of D-glucose.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U49351; AAB06943.1; -.
DR EMBL; AK052211; BAC34888.1; -.
DR EMBL; AK088481; BAC40382.1; -.
DR EMBL; BC010210; AAH10210.1; -.
DR HSSP; F01359; 2PSP.
DR MGD; MGI:95609; Gaa.
DR GO; GO:0004558; P:alpha-glucosidase activity; IMP.
DR GO; GO:0005980; P:glycogen catabolism; IMP.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
DR Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
KW SIGNAL
FT PROPEP 28 69 By similarity.
FT CHAIN 70 953 Lysosomal alpha-glucosidase.
FT DOMAIN 81 130 P-type.
FT ACT_SITE 518 518 Nucleophile (By similarity).
FT ACT_SITE 521 521 By similarity.
FT DISULFID 82 109 By similarity.
FT DISULFID 92 108 By similarity.
FT DISULFID 103 127 By similarity.
FT CARBOHYD 140 140 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 233 233 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 390 390 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 470 470 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 883 883 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 926 926 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 933 933 N-linked (GlcNAc...) (Potential).
FT CONFLICT 62 62 K -> E (in Ref. 1).
FT CONFLICT 254 254 S -> A (in Ref. 1).
FT CONFLICT 430 431 EL -> DV (in Ref. 1).
FT CONFLICT 434 434 D -> G (in Ref. 1 and 3).
FT CONFLICT 481 481 W -> C (in Ref. 1).
FT CONFLICT 615 615 G -> E (in Ref. 1).
FT CONFLICT 619 619 S -> T (in Ref. 1).
FT CONFLICT 732 732 P -> R (in Ref. 1).
FT CONFLICT 777 777 M -> V (in Ref. 1 and 3).
FT CONFLICT 871 871 R -> H (in Ref. 1).
FT CONFLICT 903 903 R -> K (in Ref. 1).
SQ SEQUENCE 953 AA; 106247 MW; 956B89685FB5FF81 CRC64;
Query Match 61.6%; Score 4016; DB 1; Length 953;
Best Local Similarity 80.7%; Pred. No. 1.6e-22;
Matches 747; Conservative 69; Mismatches 92; Indels 19; Gaps 4;
QY 320 LGYTKKHL-----QDL-----SGRISRAAEATGAHPG-----RPRVPTQCDVPP 362
DB 27 LGHLMRLMLLPQLDHSSGLWKYTPHQGKYPGLHIOEQTEQKEAPTQCDVPP 86
QY 363 NSRFDCAFDKAITQCEARGCCYIPAKQGLQAGQWPCFPFSPYSYKLENLSSSM 422
DB 87 SSRFDCAFDKISQCEARGCCYVPAGQVLKEPQIGQWPCFPFSPYSYKLENLSSSTES 146
QY 423 GYATLTTRTTTFPPKQDILRLDVMMETENLHITKDPANRRYEVLETRVHRSAPS 482
DB 147 GYATLTTRTTTFPPKQDILRLDVMMETENLHITKDPANRRYEVLETRVHRSAPS 206
QY 483 PLYSVFSEEPFGVIVRHQDGRVLLNTTVPAPFPADQFLQSLPSQYITGLAEHLSP 542
DB 207 PLYSVFSEEPFGVIVRHQDGRVLLNTTVPAPFPADQFLQSLPSQYITGLAEHLSP 266
QY 543 LMLSTWTRITLWNRDLAPTGANLYGSHPPFYLALEDGSAHGVFLNNSMMDVVLQSP 602
DB 267 LMLSTWTRITLWNRDLAPTGANLYGSHPPFYLALEDGSAHGVFLNNSMMDVVLQSP 326
QY 603 ALSWSTGILDVILFLFGPEPKSVVQVLDVVGVFPMPYWGFLGHLCRWGSSTAITRP 662

```

```
Db 327 ALTWRSTGILDVYVFLGPEPKSVWQYLDVWGVYPPMPYWGFLPFLCRMGYSSTAIVRQ 386
Qy 663 VVENMTRAFPLDQVNDLDYMSRDRFTFNKQDFRDFPAMQELHQGRRYMWIVDPAI 722
Db 387 VVENMTRTFPLDQVNDLDYMDARDFTFNQDSFADFPDMVRELHQDGRYMWIVDPAI 446
Qy 723 SSGPAGSVRPYDEGLRRGVFTINETGQLIGKVPFGSTAFDPFTNPTALAMWEDVAF 782
Db 447 SSAGPAGSVRPYDEGLRRGVFTINETGQLIGKVPFGSTAFDPFTNPTALAMWEDVAF 506
Qy 783 HQVPPDGLWIDMVEPSNIRGSDCCPNNELENPPYVGVGGTLOAATICASSHQFLS 842
Db 507 HAQVPPDGLWIDMVEPSNIRGSDCCPNNELENPPYVGVGGTLOAATICASSHQFLS 566
Qy 843 THYNLNLGLTEAIAASHALVKARTRPFISSRSTFAGHGRVAGHWTGDMVSSWQOLAS 902
Db 567 THYNLNLGLTEAIAASHALVKARTRPFISSRSTFAGHGRVAGHWTGDMVSSWQOLAS 526
Qy 903 SVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWMTOLGAFYPMFMRNHNLSLSPQEPS 962
Db 627 SVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWMTOLGAFYPMFMRNHNLSLSPQEPS 686
Qy 963 FSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQLLW 1022
Db 687 FSEPAQAMRKALTRYALLPHLYTLFHOAHVAGETVARPLFLFPKDSSTWTVDHQLLW 746
Qy 1023 GRALLITPVLOAGKARVTVGFLGTYDQVPIEALGSLPPEPPAAR-EPATHSGOMV 1081
Db 747 GRALLITPVLOAGKARVTVGFLGTYDQVPIEALGSLPPEPPAAR-EPATHSGOMV 806
Qy 1082 TLPAPLDTINVHLRAGYIIPLOQPGILTTTSSRQOPMALAVALTKGGEARGLFWDGDSL 1141
Db 807 TLEAPLDTINVHLRAGYIIPLOQPGILTTTSSRQOPMALAVALTKGGEARGLFWDGDSL 866
Qy 1142 EVLERGAYTVPLARNTVINELVRTSEGAGLQLOKTVLGATAPQOVLNSGVPSN 1201
Db 867 AVLERGAYTVPLARNTVINELVRTSEGAGLQLOKTVLGATAPQOVLNSGVPSN 926
Qy 1202 FTVSPDTKVLDICVSLMGEOPLVSW 1227
Db 927 FTVSPDTKVLDICVSLMGEOPLVSW 952

RESULT 6
ID 073632 PRELIMINARY; PRT; 873 AA.
AC 073632;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Acid alpha glucosidase.
GN Name=GAII; Synonyms=gaa2;
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98201735; PubMed=9540858; DOI=10.1016/S0925-4439(97)00092-6;
RA Kunita R., Nakabayashi O., Wu J.Y., Hagiwara Y., Mizutani M.,
RA Pennybacker M., Chen Y.T., Kikuchi T.;
RT "Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail
RT (Coturnix coturnix japonica) and the lack of its mRNA in acid maltase
RT deficient quails";
RL Biochim. Biophys. Acta 1362:269-278(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Nakabayashi O.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006754; BAA25890.2; -
DR EMBL; AB081290; BAC15596.1; -
```

```
DR HSP; Q07654; 1B9T.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
SQ SEQUENCE 873 AA; 96696 MW; CEADAFE3ACFCBD CRC64;

Query Match 50.4%; Score 3284; DB 2; Length 873;
Best Local Similarity 66.7%; Pred. No. 2e-180;
Matches 587; Conservative 117; Mismatches 156; Indels 20; Gaps 2;

Qy 348 PGRPRAVPTQCDVPPNRPDCAPDKRAITQEQCEARGCCYIPAKQGIQGAQMGPWCFFPP 407
Db 13 PSASSASASACSLPDERPDCGERLLARADCEARGCCYAPSGSGSG--GPPWCFFPL 69

Qy 408 SYPSYKLENLSSEMGYATLTTRTPFPKDLILRLDMMETENRLHFTIKDPANRRY 467
Db 70 GYRSYRADNVATAGYSARLRVVPFLPADVGTILRLDVA METESRLRFTPRDPARQY 129

Qy 468 EYPLSTPRVHSAPSLYSVEFSEBPGVIVHROLDGRVLLNTTVAPLFPADQFLQSLTS 527
Db 130 EYPMATPRVSTRAADTLXGVQLQDPFGIVVFPQDGVLLNTSVAPLFPADQFLQSLTS 189

Qy 528 LPSOYITGLAEHLSPMLSLSTWTRITLWNRDLAPTPGANLYGSHPPYLALEDGSGAHGVF 587
Db 190 LPSRIFISGLGERLAPLILDTATKTLWNRDNAPAPQVNLGSHPPYLALEDGSGAHGVF 249

Qy 588 LLNSNAMDVLQSPALSNRSTGGILDVYIFLGPEPKSVWQYLDVWGVYPPMPYWGFLG 647
Db 250 LLNSNAMDVLQSPALSNRSTGGILDVYIFLGPEPKSVWQYLDVWGVYPPMPYWGFLG 309

Qy 648 HLCRGYSGSTAITRQVVENMTFAHFLPDVQWNDLDYMSRRDFTFNKQDFRDPAMVQEL 707
Db 310 HLCRGYSGSTAITRQVVENMTFAHFLPDVQWNDLDYMSRRDFTFNKQDFRDPAMVQEL 369

Qy 708 HOGGRYMWIVDPAISSGPAGSVRPYDEGLRRGVFTINETGQLIGKVPFGSTAFDPFT 767
Db 370 HOGGRYMWIVDPAISSGPAGSVRPYDEGLRRGVFTINETGQLIGKVPFGSTAFDPFT 429

Qy 768 NPTALAMWEDVAFHFDQVFPDGLWIDMNEPSNIRGSDCCPNNELENPPYVGVGGT 827
Db 430 NPTALAMWEDVAFHFDQVFPDGLWIDMNEPSNIRGSDCCPNNELENPPYVGVGGT 489

Qy 828 LOAATICASSHQFLSTHYNLNLGLTEAIAASHALVKARTRPFISSRSTFAGHGRYAG 887
Db 490 LOAATICASSHQFLSTHYNLNLGLTEAIAASHALVKARTRPFISSRSTFAGHGRYAG 549

Qy 888 HWTGDMVSSWEOLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWMTOLGAFYPM 947
Db 550 HWTGDMVSSWEOLASSVPEILQFNLLGVPLVGADVCGFLGNTSEELCVRWMTOLGAFYPM 609

Qy 948 RHNHSLLSLPQSPYFSEFSPAQAQMRKALTRYALLPHLYTLFHOAHVAGETVARPLFL 1007
Db 610 RHNHSLLSLPQSPYFSEFSPAQAQMRKALTRYALLPHLYTLFHOAHVAGETVARPLFL 669

Qy 1008 PKDSSTWTVDHQLLWGEALLITPVLOAGKARVTVGFLGTYDQVPIEALGSLPPEPPA 1067
Db 670 PKDSSTWTVDHQLLWGEALLITPVLOAGKARVTVGFLGTYDQVPIEALGSLPPEPPA 713

Qy 1068 APREPAHSEGOMVTLAPLDTINVHLRAGYIIPLOQPGILTTTSSRQOPMALAVALTKG 1127
Db 714 -TGDSITSHKGMWILLAAPLDTINVHLRAGYIIPLOQPGILTTTSSRQOPMALAVALTKG 772

Qy 1128 EARGELFWDDGESLSEVLERGAYTVQVIFLARNTVINELVRTSEGAGLQLOKTVLGAT 1187
Db 773 EARGELFWDDGESLSEVLERGAYTVQVIFLARNTVINELVRTSEGAGLQLOKTVLGAT 832

Qy 1188 APQOVLNSGVPSNFTYSPDTKVLDICVSLMGEOPLVSW 1227
Db 1188 APQOVLNSGVPSNFTYSPDTKVLDICVSLMGEOPLVSW 1227
```

```

Db 833 APQVLANGIPVEDFSYRSDDTQVLRVSVSLPMWQFVVPW 872
RESULT 7
ID 073626 PRELIMINARY; PRT; 932 AA.
AC 073626;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE alpha glucosidase.
GN Name=GAAL; Synonyms=Gaal;
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98201735; PubMed=9540858; DOI=10.1016/S0925-4439(97)00092-6;
RA Kunita R., Nakabayashi O., Wu J.Y., Hagiwara Y., Mizutani M.,
RA Pennybacker M., Chen Y.T., Kikuchi T.;
RT "Molecular cloning of acid alpha-glucosidase cDNA of Japanese quail
RT (Coturnix coturnix japonica) and the lack of its mRNA in acid maltase
RT deficient quails."
RL Biochim. Biophys. Acta 1362:269-278(1997).
RV [2]
RP SEQUENCE FROM N.A.
RA Nakabayashi O.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB000967; BAA25884.1; -
DR EMBL; AB081289; BAC15595.1; -
DR HSSP; Q07654; 189T.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000322; Glyco hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 1.
DR Pfam; PF00088; Trefoil; 1.
DR SMART; SM00018; PD; 1.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
SQ SEQUENCE 932 AA; 104689 MW; B62E182F03DE3F61 CRC64;
Query Match 39.5%; Score 2573.5; DB 2; Length 932;
Best Local Similarity 53.0%; Pred.No.1.8e-139;
Matches 472; Conservative 144; Mismatches 248; Indels 27; Gaps 9;
QY 341 EAETGAHGRPRAVPTQCD--VPNSRFDCAPDK--AITQCEARGCCYIPAKOGLQGAQ 397
Db 64 QVSRGWGHPAP---PAQCHLVESHRYDCYPERNVVVTQELCESGCCFIQTLPVAV-GGK 119
QY 398 MGQPCWFFPPSPYSKLENSSEMGTATLRTTTPFPKDLTLRLDVMETENRLHF 457
Db 120 RGVPCWCFPPDPSPYSVQSLNQTGLVRRERKAYPKDIQMLRMDVETQNTLHI 179
QY 458 TIKDPAHREYVPLETPRVHRSAPLSYSEPEPFGVIVHRLDGRVLLNTTAPLPF 517
Db 180 KITDAANRYEVLPRVTRKAEENIYSLSEISQDPFGVLLRRQGTGTVLLNTTAPLPF 239
QY 518 ADQFLQSLTSLRSQITGLAEHLSPMLSTSTRTITLWNRDLAPTGANLYGSHFFYLAL 577
Db 240 ADQFLQISITLRSFYGLGEHRSTLLHSLDWNLTFLWADVAPTESFNLYGAHPFYLAM 299
QY 578 EGGSHGAVFLNSNAMDVVLPSPALSWRSCTGILDVIFLGPSPKSVVQQYLDVVGYP 637
Db 300 EGGDAHGAVFLNSNAMEVALQFAPGLTWRITGGVLDYIFLGPDPNMVVIQQYQVEIGFP 359
QY 638 FMPYVWGLGFHLRCWCGYSTAITQVVENNTRAHPPLDVQWDLVDYMDSRDFTFNKGF 697
Db 360 AMPPLWALGFHLRCWCGYGSNETWQTAKARNFQIPQDAQWDLVDYMDGRDFTFDQPF 419

```

```

QY 698 RDPFAMVQELHQGRRYMMIVDPAISSSPAGSYRYPYDEGLRGRVFIETNETQOPLIGKW 757
Db 420 ASPLSLVEDLHKHGQHYIILDPGISSSTSPRSYWPFDGLRGLFLNTTQGTLLQGVW 479
QY 758 PGSTAPDPFTPTALAWEDMVAEPHDQVPPDGLMIDMNEPSNPIRGSDGCGNNELENP 817
Db 480 PGYTAYDPDSNTDTHQWLENLQRFHTVPPFDGLMIDMNEPSNPMDSGEGCPGELDSP 539
QY 818 PVPGVVGTTLQAATTCASSHQFLSTHYNLNLHLYGLTEAIAASHRALVKARGTRPPVIS 877
Db 540 PYPYAVLGNLSUTAKTVCASAEQNASVHYNLNLHLYGLKEAEATASALIRIRGRPPVIS 599
QY 878 TFAGHGRYAGHWITGDVMSWEQLASSVPEILQFNLLGVPLVGVADYCGFLGNTSEELCV 937
Db 600 TFPQSGRYSGHWLGDNRSQWDMYYSIFGMLSFSFLGIPLVGADICGFSGSTSEELCTRW 659
QY 938 TOLGAFYFPMNHNLSLSPQEPYSPFAQQAOMRKALTLRYALLPHLYTLFPHQAHVAGE 997
Db 660 MOLGAFYFPMNHNQNEKAQDPTAFSPSARTAMKDALLTRYSLLPFLYTLFPHRAHQE 719
QY 998 TVARPLFLFEFPKDSSTWTDVHQLLWGEALLITPVLQAGKAEVTPFPPLGTWYDLQVPI 1057
Db 720 TVARPLFEFPWDVATYGLDRQFLMGQSLVTPVLEPGADSVLGYFPQGVWYDFYT---- 775
QY 1058 ALGSLPPPPAAPREPAIHSEGOVTLPAPLDTINVHLRAGYIIPLOQPLGTTTESRQPM 1117
Db 776 --GS-----SVNSSGEMLKLSAPLDHLNLHREGSILPTQKPGITSKATRGNDL 822
QY 1118 ALAVALTGGEGARGELFWDDGESLEVLERGAYTVQVIFLARNTIYNELVRVTSEAGLOL 1177
Db 823 HLIVALTSTRATAWGDLFWDDGESLDTFEGNSYLVFNATENIFTSNVLHASTEATDVTI 882
QY 1178 QKVTVLGVATAPQOVLNGVPSVNFYSPDTKVLDC--VSLLMGBOFLVSW 1227
Db 883 DAVSYFGVQEPSPKVLDDG--QEKPSYL--DNQVLTSLGLVLSQGSFLQW 931
RESULT 8
MGA HUMAN
ID MGA_HUMAN STANDARD; PRT; 1856 AA.
AC 04351;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Maltase-glucosylase, intestinal (Includes: Maltase (EC 3.2.1.20)
DE (Alpha-glucosidase); Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosidase)].
GN Name=MGA; Synonyms=MGA, MGAML;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Small intestine;
RX MEDLINE=98112863; PubMed=9446624; DOI=10.1074/jbc.273.5.3076;
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RT "Human small intestinal maltase-glucoamylase cDNA cloning. Homology to
RT sucrase-isomaltase."
RL J. Biol. Chem. 273:3076-3081(1998).
RV [2]
RN REVISIONS TO 776; 1049; 1100; 1541; 1612 AND 1811.
RA Nichols B.L., Eldering J.A., Avery S.E., Hahn D., Quaroni A.,
RA Sterchi E.E.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RV [3]
RP CHARACTERIZATION.
RC TISSUE=Small intestine mucosa;
RX MEDLINE=89066802; PubMed=3143729;
RA Naim H.Y., Sterchi E.E., Lentze M.J.;
RT "Structure, biosynthesis, and glycosylation of human small intestinal
RT maltase-glucoamylase."

```


Db 729 AHSRGDTVARPLLEHFEYEDNSTWDHQQFLWGPGLLITPVLDEGAEKVMAYVPDVAWYDY 788
Qy 1052 QVPIEALGLSPPPAAPREPAHSEGOVWTLPAPLDTINVHLRAGYIIPLOQPGLTTE 1111
Db 789 ET-----GS-----QVRWKQKQVEMELFODKIGLHRLRGYIPFTQOPNTTLLA 831
Qy 1112 SRQQMALAVALTKGGEALFWDGSLVLERGAYTQVIF-LARNNTIYNELVRVTS 1170
Db 832 SRKNPLGLIILADENKEAKGELFWDGGETKDTVANKVYLCEFSVTQNRLEVNISQSTVK 891
Qy 1171 EGAGLOLQKTVLGVATAPQ-QVLSNGVPVS---NFTYSPDTKVLIDIC-VSLLMGEQFLV 1225
Db 892 DPNNAFNEIKILGTEPNSVTVKNGVPSQTSPTVTYDNLKVAIITDIDLLGEAYTV 951
Qy 1226 SW 1227
Db 952 EW 953

RESULT 9
SUIS_HUMAN
ID SUIS_HUMAN STANDARD; PRT; 1826 AA.
AC P14410;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN Name=SI;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=92359963; PubMed=1353958;
RA Chantret I., Lacasa M., Chevalier G., Ruf J., Islam I., Mantei N.,
RA Edwards Y., Swallow D., Rousset M.;
RT "Sequence of the complete cDNA and the 5' structure of the human
RT sucrase-isomaltase gene. Possible homology with a yeast
RT glucosylase.";
RL Biochem. J. 285:915-923 (1992).
RN [2]
RP SEQUENCE OF 1-677 FROM N.A.
RX MEDLINE=88112852; PubMed=2962903; DOI=10.1016/0378-1119(87)90181-8;
RA Green F., Edwards Y., Hauri H.-P., Povey S., Ho M.W., Pinto M.,
RA Swallow D.;
RT "Isolation of a cDNA probe for a human jejunal brush-border hydrolase,
RT sucrase-isomaltase, and assignment of the gene locus to chromosome
RT 3.";
RL Gene 57:101-110 (1987).
RN [3]
RP SEQUENCE OF 1-19 AND 1007-1023, AND TISSUE SPECIFICITY.
RX PubMed=1677636;
RA Gervel J.P., Ferrero A., Chambrud L., Rigal A., Bonicel J.,
RA Maroux S.;
RT "Expression of sucrase-isomaltase and dipeptidylpeptidase IV in human
RT small intestine and colon.";
RL Gastroenterology 101:618-625 (1991).
RN [4]
RP VARIANT DISACCHARIDE INTOLERANCE I PRO-1097.
RX MEDLINE=96189940; PubMed=8609217;
RA Ouwendijk J., Moolenaar C.E.C., Peters W.J., Hollenberg C.P.,
RA Giesel L.A., Franssen J.A.W., Naim H.Y.;
RT "Congenital sucrase-isomaltase deficiency: identification of a
RT glutamine to proline substitution that leads to a transport block of
RT sucrase-isomaltase in a pre-Golgi compartment.";
RL J. Clin. Invest. 97:633-641 (1996).
CC -!- FUNCTION: Plays an important role in the final stage of
CC carbohydrate digestion.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-

D-glucosidase-type action.
-!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
in some oligosaccharides produced from starch and glycogen by
alpha-amylase, and in isomaltose.
-!- SUBUNIT: The resulting sucrase and isomaltase subunits stay
associated with one another in a complex by non-covalent linkages.
-!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
-!- TISSUE SPECIFICITY: Expressed in the poorly differentiated crypt
cells of the small intestine as well as in the mature villous
cells. Expressed at very low levels in the colon.
-!- PTM: The precursor is proteolytically cleaved when exposed to
pancreatic proteases in the intestinal lumen.
-!- PTM: Sulfated (By similarity).
-!- DISEASE: Defects in SI are the cause of disaccharide intolerance I
[MIM:222900].
-!- MISCELLANEOUS: There is a high degree of homology between the
isomaltase and sucrase portions (41% of amino acid identity)
indicating that this protein is evolved by partial gene
duplication.
-!- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.
-!- SIMILARITY: Contains 1 P-type (trefoil) domain.
-!- SIMILARITY: Contains 1 P-type (trefoil) domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; X63597; CAA45140.1; -
DR EMBL; M22616; AAA60551.1; ALT_SEQ.
DR PIR; S36082; UUUU.
DR Genew; HGNC:10856; SI.
DR MIM; 222900; -
DR GO; GO:0005903; C:brush border; TAS.
DR GO; GO:0005794; C:Golgi apparatus; TAS.
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 2.
DR Pfam; PF00088; Trefoil; 2.
DR SMART; SM00018; PD; 2.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 1.
DR PROSITE; PS00025; P_TREFOIL; 1.
KW Direct protein sequencing; Disease mutation; Glycoprotein;
KW Glucosidase; Hydrolase; Multifunctional enzyme; Repeat; Signal-anchor;
KW Sulfation; Transmembrane.
FT INIT_MET 0
FT CHAIN 1 1826 Sucrase-isomaltase, intestinal.
FT CHAIN 1 1006 Isomaltase.
FT CHAIN 1007 1826 Sucrase.
FT DOMAIN 1 11 Cytoplasmic.
FT TRANSMEM 12 31 Signal-anchor for type II membrane
protein.
FT DOMAIN 32 1826 Lumenal.
FT DOMAIN 61 108 P-type.
FT DOMAIN 109 1006 Isomaltase.
FT DOMAIN 1007 1826 Sucrase.
FT DOMAIN 42 59 Ser/Thr-rich.
FT ACT_SITE 504 504 Nucleophile.
FT ACT_SITE 507 507 By similarity.
FT ACT_SITE 1393 1393 Nucleophile.
FT ACT_SITE 1396 1396 By similarity.
FT ACT_SITE 1499 1499 Proton donor (By similarity).
FT DISULFID 62 93 By similarity.
FT DISULFID 76 92 By similarity.
FT DISULFID 87 105 By similarity.
FT MOD_RES 236 236 Sulfotyrosine (potential).
FT MOD_RES 238 238 Sulfotyrosine (potential).
FT MOD_RES 390 390 Sulfotyrosine (potential).
FT MOD_RES 399 399 Sulfotyrosine (potential).
FT MOD_RES 666 666 Sulfotyrosine (potential).


```
FT MOD_RES 762 762 Sulfoxyrosine (Potential).
FT MOD_RES 764 764 Sulfoxyrosine (Potential).
FT CARBOHYD 98 98 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 436 436 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 454 454 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 822 822 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 854 854 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 903 903 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 925 925 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1234 1234 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1302 1302 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1339 1339 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1353 1353 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1402 1402 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1534 1534 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1571 1571 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1674 1674 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1747 1747 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1762 1762 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1814 1814 N-linked (GlcNAc. .) (Potential).
FT VARIANT 1097 1097 Q -> P (in disaccharide intolerance I;
exhibits intracellular accumulation of
mannose-rich SI in the Golgi).
/FTId=VAR_007854.
V -> F (in Ref. 3).
Missing (in Ref. 2).
S -> E (in Ref. 3).
V -> T (in Ref. 3).
SEQUENCE 1826 AA; 209272 MW; 377E4B66DFCF9C8E CRC64;

Query Match 28.8%; Score 1877.5; DB 1; Length 1826;
Best Local Similarity 41.7%; Pred. No. 5.9e-99;
Matches 382; Conservative 157; Mismatches 317; Indels 59; Gaps 19;

QY 340 AEAECAHGRPRAVTQCDVPSNRPDCADPAKAITQCEARGCCYIPAKQGLQAGWG 399
DB 48 ATTRVTNPSDSKCNVLNDPVNVNINCPIDPEQFTGICAQGCCWRPNWDSL----- 101
QY 400 QWCFPPSPVSKLENLSSEMGYATLTR-TPTFFPKDILTLRDVMMETENLHFT 458
DB 102 IPWCFVDNH-GYNVQDMTTSIGVEAKLRIPSPTLFGNDINSVLTQNTQNPFRFK 160
QY 459 IKDPAHRYEVPLETRVHS--RAPSLYSVEFSEBPPGVVHRQDGRVLLNTTVAFLP 516
DB 161 ITDPNRRYEVHPQYKEFTGPTVSDLYDKVAQNPFISQVIRKSGKGLFTDSIGPLV 220
QY 517 FADQFLQSLSPQVITGLAEHL-SPLMLSTSWTRITLWNRDLAP-TPGANLYGSHPFY 574
DB 221 YSDQYLQISARLPSDIYIGGEQVHKRFRHDLWSKWTPIFTRDQLPCDNNNLYGHQTF 280
QY 575 LALED-GGSAHGVPFLNSAMDVLPSPALSWRSRSTGGILDVYIFLGPBKSYYQQYLDV 633
DB 281 MCIEDTSKSGFGVFLNSNMEIFQTPVTVRVTVGGILDVFIYLLGDTPEQVVQQYQL 340
QY 634 VGYPFPPYWGFLHLCRMGYSSAITRQVVENMTRAHFPDLDVQWDLDDMSRRDFTFN 693
DB 341 VGLPAMPAYNWLGFQLSRWYKSLDVVKEVVRNREAGIPFDQVTDIDVMEKDKFTVD 400
QY 694 KDGFRDPPAMQBLHGGRRYMMVDPATSSSGPAG--SYRPYDEGLRGVFTNTEGQ- 750
DB 401 QVAFNLGPQVQDLHDHGQVILDPATISIGRRANGTTVATYERGNTQHWLNESDGST 460
QY 751 PLIGKVPKSTAPDFTNPTALAWEDMVAERHDQVDFDGLWDMNEPSNFIKSGSDGCP 810
DB 461 PIIGEVWPGUTVVPDFTNPDNDWANECSIFHQEVQYDGLWDMNEVSFIOGSKYGCN 520
QY 811 NNELEPPYVPGVGGTQQAATCASHQPLSTHYNLHNLGLTEATASHRAVLKA-RGT 869
DB 521 VNKLNYPPPTDILDKMLSKTICMDAVQWQKQYDVHSLYGYSMALTEQAQVKQFPNK 580
QY 870 RPFVISRSTAGHVRAGHTGWDWSWEQLASVPEILQFNLGLVPLVGADVCGFLGNT 929
DB 581 RSPILTRSTAGSGRAAAHLWGDNTASWEQMEWSITGMLEFSLFGIPLVGADICGFAET 640
```

```
QY 930 SBEELVRWTQLCAPYFPMNHNLSLSLPQBPYSFSEPA-----QQAMRKALTURYAL 981
DB 641 TBEELCRMMQLCAPYFPMNHNLSLSLPQBPYSFSEPA-----DGYEHQDPAFFGQNSLVKSRQYLTIRYTL 694
QY 982 LPHLYTLPHQAVAGETVARPLPLEFPKDSSTWTVDDHQLLWGEALLITPVLQAGKAEVIG 1041
DB 695 LPFLYTLFYKAHVFGETVARPVLHFEYDNTSWIEDTEFLMGPALDITPVLKQAGDVSA 754
QY 1042 YPLGTYDLOTVPIEALGSLPAPPAAPRAIPHSEGOVVTLPAPLDTINVHLRAGYIIP 1101
DB 755 YIPDAIWDYES-----GAKRP-----WRKQVDMYLPADKIGLHLRGYIIP 797
QY 1102 LQPGGLTTTTSROQPWALAVLTKGGEARGELFWDDGESLEVLERGAYTQVIFLARNTI 1161
DB 798 IQEPDVTTTTARKNPLGLIVALGENTAKGDFDWDGETKDTQNGNYILYTPSVSNNTL 857
QY 1162 VNELRYVTS--EGAGLQLOKTVVLGV--ATAPOQVLSNGVPV--SNTYSPTDKVLDI 1213
DB 858 --DIVCTHSSYQEGTTLAFQTVKILGLTSDVTEVRVAENNQPMNAHSNFTYDASNQVLLI 915
QY 1214 C-VSLLMGBOQLVSW 1227
DB 916 ADLKNLGNFNSVQW 930

RESULT 10
SUIS_RABIT STANDARD; PRT; 1826 AA.
AC P07768;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN NamesSI;
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=86245068; PubMed=3755079; DOI=10.1016/0092-8674(86)90739-7;
RA Hunziker W., Spiess M., Semenza G., Lodish H.F.;
RT "The sucrase-isomaltase complex: primary structure, membrane-
RT orientation, and evolution of a stalked, intrinsic brush border
RT protein.";
RL Cell 46:227-234 (1986).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-37 AND 1007-1014.
RX MEDLINE=83105704; PubMed=7152027; DOI=10.1016/0014-5793(82)80833-8;
RA Sjoestrom H., Noren O., Christiansen L.A., Wacker H., Spiess M.,
RA Bigler-Meier B., Rickli E.E., Semenza G.;
RT "N-terminal sequences of pig intestinal sucrase-isomaltase and pro-
RT sucrase-isomaltase. Implications for the biosynthesis and membrane
RT insertion of pro-sucrase-isomaltase.";
RL FEBS Lett. 148:321-325 (1982).
CC -I- FUNCTION: Plays an important role in the final stage of
carbohydrate digestion.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-
D-glucosidase-type action.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
in some oligosaccharides produced from starch and glycogen by
alpha-amylase, and in isomaltose.
CC -I- SUBUNIT: The resulting sucrase and isomaltase subunits stay
associated with one another in a complex by non-covalent linkages.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -I- PTM: The precursor is proteolytically cleaved when exposed to
pancreatic proteases in the intestinal lumen.
CC -I- PTM: N- and O-glycosylated.
CC -I- PTM: Sulfated (By similarity).
CC -I- MISCELLANEOUS: There is a high degree of homology between the
isomaltase and sucrase portions (41% of amino acid identity)
```

indicating that this protein is evolved by partial gene duplication.
 -1- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.
 -1- SIMILARITY: Contains 1 p-type (trefoil) domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M14046; AAA31459.1; -
 DR InterPro: IPR000322; Glyco_hydro_31.
 DR InterPro: IPR000519; P_trefoil.
 DR Pfam; PF01055; Glyco_hydro_31; 2.
 DR Pfam; PF00088; Trefoil; 2.
 DR SMART; SM00018; PD; 2.
 DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
 DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; 2.
 DR PROSITE; PS00025; P_TREFOIL; 1.
 KW Direct protein sequencing; Glycoprotein; Glycosidase; Hydrolase;
 KW Multifunctional enzyme; Repeat; Signal-anchor; Sulfation;
 KW Transmembrane.

INIT MET 0
 FT CHAIN 1 1826 Sucrase-isomaltase, intestinal.
 FT CHAIN 1 1006 Isomaltase.
 FT CHAIN 1007 1826 Sucrase.
 FT DOMAIN 1 11 Cytoplasmic.
 FT TRANSMEM 12 31 Signal-anchor for type II membrane protein.
 FT DOMAIN 32 1826 Lumenal.
 FT DOMAIN 42 59 Ser/Thr-rich.
 FT DOMAIN 61 108 P-type.
 FT DOMAIN 109 1006 Isomaltase.
 FT DOMAIN 1007 1826 Sucrase.
 FT ACT SITE 504 507 Nucleophile.
 FT ACT SITE 1393 1393 By similarity.
 FT ACT SITE 1396 1396 Nucleophile.
 FT ACT SITE 1499 1499 By similarity.
 FT DISULFID 62 93 Proton donor (By similarity).
 FT DISULFID 76 92 By similarity.
 FT DISULFID 87 105 By similarity.
 FT MOD_RES 390 390 Sulfotyrosine (Potential).
 FT MOD_RES 399 399 Sulfotyrosine (Potential).
 FT MOD_RES 1381 1381 Sulfotyrosine (Potential).
 FT MOD_RES 1384 1384 Sulfotyrosine (Potential).
 FT CARBOHYD 41 41 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 98 98 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 454 454 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 858 858 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 895 895 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 903 903 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1001 1001 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1234 1234 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1302 1302 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1324 1324 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1339 1339 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1353 1353 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1367 1367 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1402 1402 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1534 1534 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1571 1571 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1747 1747 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1762 1762 N-linked (GlcNAc..) (Potential).
 FT CARBOHYD 1798 1798 N-linked (GlcNAc..) (Potential).
 SQ SEQUENCE 1826 AA; 210008 MW; 6840D03955A45BB5 CRC64;

Query Match 28.8%; Score 1875.5; DB 1; Length 1826;
 Best Local Similarity 42.0%; Pred. No. 7.7e-99;
 Matches 377; Conservative 155; Mismatches 312; Indels 53; Gaps 19;

QY	355	PTQCDVPPNSRFDCAADKAITQEQEARGCCYIPAKQGLQGAQMGPWCFFPSPSYKL	414
DB	63	PSELNEVNERINCIPESQPTQAIQAQNCWCWRPNNS-----DIPMCFVDNH-GYNV	115
QY	415	ENLSSENGYATLTR-TTPFPFKDILTLRLDVMMETENRLHFTIKDPANRYEVLB-	472
DB	116	EGMTTSTGLEARLNKSTPTLFGNDINNVLTTESQTANRLRFLKLTDPNNKRYEVP	175
QY	473	TPRVHSRAPSLYSVESESPGVIVHRLQDGRVLLNTVAPLFFAQLFOLSLSLPSQ	531
DB	176	VTEFAGPATETLYDVQVTENPFISKIRKSNRILFDSSIGPLVYSQYLQISRLPSE	235
QY	532	YITGLAHL-SPLMLSTSWTRITLWNRDL-APTGANLYGSHPPFYLALED-CGSAH	588
DB	236	YMGFGVHVKFRHDLWKTPITRQHTDDNNNLYGHQTFEMCIEDTTGKSGVFL	295
QY	589	LNSNADVVLPSPALSWRSTGGILDVIFLGPESKVVQVLDVVGVVFPMPYVGLGPH	648
DB	296	MNSNAMEIFIOPTPIVTVRVIGGILDVIFLGDTPQVQVQYQELIGRPAMPAYMS	355
QY	649	LCRWGYSSTAIQVVENMTRAHFPLDVQWDLVDMSRRDFTENKDGFRDPAMVQEL	708
DB	356	LSRWYNSLDVVVRRNRREALIPFTQVSDIDIMEDKDKFTYDRVANGLPDFVQDL	415
QY	709	QGGREYMMIVDPAIS-----SSGPAGSYRYPDEGLRRGVFITNETG-QPLIG	763
DB	416	DHGQKYVILDPAISINRRASGEA--YESYDRGNAQVWVNESDGTTPIVGEVMP	473
QY	764	PDFTNPTLAWWEDMVAFHDQVPPDGLWIDMNEPSNFIKSGEDCGPNNELENP	823
DB	474	PDFTSPNCIEWANEENIFHOEVNYDGLWIDMNEVSVFQSGNKGCDNTLNYPI	533
QY	824	VGGTLQAATICASSHOFSLTHYNLNLGLTEATASHRALVKA-RGTSPFVLSR	882
DB	534	VDKLMYSKTLCDMSVQYWGKQYDVHSLGYSNATIAERAVRFPNKRKSFILTR	593
QY	883	GRYAGHWGDMVSSWEQELASSVPEILOFNLGVLGVADVCGFLGNTSEELCVRT	942
DB	594	GRHAAHLGDNTATWEQWMSITGNLEFGLGMPVGVADICGFLAETTELCCR	653
QY	943	FYPFRNHNLSLSLPQEPYSPSEPA--QOAMRKALTRYALLPHLYTYLPHQAH	1000
DB	654	FYPFRNHNADGFHQDPAPFGQDLSLVKSSRHYLNIRYLLPFLYTLFYKAHA	713
QY	1001	RPLFLFEPKDSSTVTDVHQLLWGEALLTPVLQACKAEVTVGFLGTWDLQTV	1060
DB	714	RPVLHEFDVNSWVEDREFLWGPALLITPVLVTQAEVTSVATPDAVWDYDET	767
QY	1061	SLPPPPAAPREPPIHSEGWVTLPAPLDTINVHLRAGYIIPQGPGLTTESRQ	1120
DB	768	AKRP-----WRQVEMSLPADKIGLHLAGGYIIPQPAVITTSRMNPLGLI	816
QY	1121	VALTKGBARGELFWDDGESLEVLRGAYTVQVIFLARNN-----TIVNEL	1176
DB	817	IALNDNTAVGDFFWDDGCTKDTQNDVNYLYTFAVSNNNLNIITCTHELY---	873
QY	1177	LQVTVLGVATAPQ--VLSNGVPV---SNFTYSPDTKVLDI-CVSLLMGEOLF	1227
DB	874	FQTIKILGVTVTQVTVVAENNSQMSHNSFTYPSNQVLLIENLNLGRNFRVQ	930
RESULT 11			
ID	Q8TE24	PRELIMINARY;	PRT; 1734 AA.
AC	Q8TE24;		
DT	01-JUN-2002	(TReMBLrel. 21, Created)	
DT	01-JUN-2002	(TReMBLrel. 21, Last sequence update)	
DT	01-JUN-2003	(TReMBLrel. 24, Last annotation update)	
DE	Maltase-glucoamylase (Fragment).		
GN	Name=MGAM;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22457205; PubMed=12547908; DOI=10.1073/pnas.0237170100;
 RA Nichols B.L., Avery S., Sen P., Swallow D.M., Hahn D., Sterchi E.;
 RT "The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase
 with complementary starch digestion activities.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1432-1437(2003).
 RN SEQUENCE FROM N.A.
 RP Nichols B.L. Jr., Avery S.E., Sen P., Swallow D.M., Hahn D.,
 RA Sterchi E.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF432202; AAL83560.1; JOINED.
 DR EMBL; AF432187; AAL83560.1; JOINED.
 DR EMBL; AF432188; AAL83560.1; JOINED.
 DR EMBL; AF432189; AAL83560.1; JOINED.
 DR EMBL; AF432190; AAL83560.1; JOINED.
 DR EMBL; AF432191; AAL83560.1; JOINED.
 DR EMBL; AF432192; AAL83560.1; JOINED.
 DR EMBL; AF432193; AAL83560.1; JOINED.
 DR EMBL; AF432194; AAL83560.1; JOINED.
 DR EMBL; AF432195; AAL83560.1; JOINED.
 DR EMBL; AF432196; AAL83560.1; JOINED.
 DR EMBL; AF432197; AAL83560.1; JOINED.
 DR EMBL; AF432198; AAL83560.1; JOINED.
 DR EMBL; AF432199; AAL83560.1; JOINED.
 DR EMBL; AF432200; AAL83560.1; JOINED.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR000322; Glyco_hydro_31.
 DR Pfam; PF01055; Glyco_hydro_31; 2.
 DR Pfam; PF00088; Trefol; 1.
 DR SMART; SM00018; PD; 1.
 DR PROSITE; PS00129; GLYCOSYL_HYDROL_P31_1; 2.
 FT NON_TER 1
 SQ SEQUENCE 1734 AA; 196946 MW; BAF70D11BEDDBA7E CRC64;
 Query Match 28.1%; Score 1830; DB 2; Length 1734;
 Best Local Similarity 44.4%; Pred. No. 3e-96;
 Matches 376; Conservative 129; Mismatches 303; Indels 38; Gaps 18;
 QY 401 PNCFFPPSPYSKLE-NLSSSEMGYATITRTPT--PPFKDILRLDMMETENLRHF 457
 DB 5 PNCYSKNH-SYHVEGLNVTNAGFTARL-KNLPSPVFGSNVDNVLTAETQSNRHF 62
 QY 458 TKDPAIRRYEVPLETPRVHS--RAPSPYSVEFSEEPFGVIVHROLDGRVLLNTTVA 515
 DB 63 KLTQDTNRFEPHVEHVQSFSGNAASLTQVEISRQPSIKVTRSNRNVLPDSSIGPL 122
 QY 516 PFADQFLQSTSPSQYITGLAEHL-SPLMLSTWTRITLWNRDLAPT-PGANLYGSHPF 573
 DB 123 LFADQFLQSTRLPSTNVTYGLGEHVQVHRHDMNWTWFIENRDTTPNGNGLYGAQTF 182
 QY 574 YLALIED-GGSAHGAVFLNENAMDVVLQSPALSWRSTGGILDYVIFLGPPEKSVVQOYLD 632
 DB 183 FLCLEDASGLSGVFLMNSNMEVLPAPAITRTIGGILDYVFLGNTPEQVQYXLE 242
 QY 633 VVGYPMPYPYWGFLHLCRWGYSSTAITQVVENMTTRAHPDLDVQNDLDYDMSRDRFTF 692
 DB 243 LIGRPALPSYALGFLHLSRYEYGTLDNMREVVERNRAQLPYDVQHADIDYMDERRFTY 302
 QY 693 NKDGRDPFAMVQELHOGGRRYIMVDPAIS--SSGPAQSVYRVDGLRGVFTNETG-Q 750
 DB 303 DSVDFKGFPEFVELHNNQOKLVIIVDPAINSSSSKPGYPVDRGSDVKIWNSSDGV 362
 QY 751 PLIGKVPWGSTAPPDFTNTPTALAWMDMVAEFHQVDFPDGLMDNPNESNFIKSGSDGCP 810
 DB 363 PLIGVWPQCTVFPDFTNTNCAVWTKPELFHNVQVEFDGIDMNMVENFVDSVSGCS 422
 QY 811 NNELENPPYVPGVGGTLOQAATCASSHQFLSTHYNLHNLGLTEAIAASHRALVKA-RGT 869
 Db 423 TNNLANPPPTPRILDGYFCCKLMDAVQHWKGVDIHNLGYSMATAAABAAKTVFPNK 482
 QY 870 RPFVSRSTFAGHRYAGHTGDVWSSWEQLASSVPEILQFNLLGVPLVGVADVCGFLGNT 929
 Db 483 RSFILTRSTFAGSKFAHNLGDNTATWDDLWSIPGVLEFNLFGIPMGVDPICGALDT 542
 QY 930 SBEICVRWTLQCAFYPFWRNHNLSLSPQEPVSFSEPA--QOAMRKALTLRVALLPPLYT 987
 Db 543 PEELCRNMQLGAFYFPRNHNQGYKQDQDPASFGADSLNLSRRHLYNIRVTLPLPLYT 602
 QY 988 LFHQAHVAGETVARPLPLEFPKDSSTWTVHQLLAGEALLITPVLOQAKAENVGYEPLGT 1047
 Db 603 LFFRAHSGDITVARPLLHFEYEDNSTWVHQFLWGPGLLITPVLDGEAKKWAYVDPDAV 662
 QY 1048 WYDLQVTPTEALGSLPPPPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYIIPLOGPGL 1107
 Db 663 WYDYET-----GS-----QVRWRKQKVEMLPGDKLGLHLRGYIFPTQPNT 705
 QY 1108 TTESRQOPMALAVALTGGEGARGELFWDGSGSLEVLERGATVQVIFLARNNTIVNELVR 1167
 Db 706 TTLASRKNPLGLIIALDENKEAKGELFWDGSGTAKTKDTAKKVVLLCEFSVTQNHLEVTISQ 765
 QY 1168 VT-SEGAGLQKQVTLGVATAPQ-QVLSNGVPVS---NFTYSPDTKVLDC-VSLLMGE 1221
 Db 766 STYKDPNNLAFNEIKILGMBSNVTVXHGVPSTPTPTVTVDSNLKVAITDIDLLGE 825
 QY 1222 QFLVSW 1227
 Db 826 AYTVEM 831
 ID SUIS RAT STANDARD; PRT; 1840 AA.
 AC P23739;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
 DE Isomaltase (EC 3.2.1.10)].
 GN Name=Si;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Intestine;
 RX MEDLINE=95121929; PubMed=7821806; DOI=10.1016/0378-1119(94)90452-9;
 RA Chandrasena G., Osterholm D.E., Sunitha I., Henning S.J.;
 RT "Cloning and sequencing of a full-length rat sucrase-isomaltase-
 encoding cDNA";
 RL Gene 150:355-360(1994).
 RN [2]
 RP SEQUENCE OF 86-361 FROM N.A.
 RC STRAIN=Fischer 344; TISSUE=Intestine;
 RX MEDLINE=91097578; PubMed=2268340;
 RA Traber P.G.;
 RT "Regulation of sucrase-isomaltase gene expression along the crypt-
 villus axis of rat small intestine.";
 RL Biochem. Biophys. Res. Commun. 173:765-773(1990).
 RN [3]
 RP SEQUENCE OF 732-1372 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Duodenum;
 RX MEDLINE=90381315; PubMed=2400788; DOI=10.1016/0167-4781(90)90121-H;
 RA Broyart J.-P., Hugot J.-P., Perret C., Porteu A.;
 RT "Molecular cloning and characterization of a rat intestinal sucrase-
 isomaltase cDNA. Regulation of sucrase-isomaltase gene expression by
 sucrose feeding.";
 RL Biochim. Biophys. Acta 1087:61-67(1990).
 RN [4]
 RP SEQUENCE OF N-TERMINUS OF ISOMALTASE AND SUCRASE.

RX MEDLINE=82167542; PubMed=6802834;
 RA Hauri H.-P., Wacker H., Rickli E.E., Bigler-Meier B., Quaroni A.,
 RA Semenza G.;
 RT "Biosynthesis of sucrose-isomaltase. Purification and NH2-terminal
 RT amino acid sequence of the rat sucrose-isomaltase precursor (pro-
 RT sucrose-isomaltase) from fetal intestinal transplants.";
 RL J. Biol. Chem. 257:4522-4528 (1982).
 CC -!- FUNCTION: Plays an important role in the final stage of
 CC carbohydrate digestion.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-
 CC D-glucosidase-type action.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
 CC in some oligosaccharides produced from starch and glycogen by
 CC alpha-amylase, and in isomaltose.
 CC -!- SUBUNIT: The resulting sucrose and isomaltase subunits stay
 CC associated with one another in a complex by non-covalent linkages.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
 CC -!- PTM: The precursor is proteolytically cleaved when exposed to
 CC pancreatic proteases in the intestinal lumen.
 CC -!- PTM: Sulfated (By similarity).
 CC -!- MISCELLANEOUS: There is a high degree of homology between the
 CC isomaltase and sucrose portions (41% of amino acid identity)
 CC indicating that this protein is evolved by partial gene
 CC duplication.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.
 CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L25926; AAA5097.1; -;
 DR EMBL; M62889; AAA42144.1; -;
 DR EMBL; X15546; CAA33552.1; -;
 DR PIR; T10799; T10799.
 DR RGD; 3675; Si.
 DR InterPro; IPR000322; Glyco_hydro_31.
 DR InterPro; IPR000519; P_trefoil.
 DR Pfam; PF01055; Glyco_hydro_31; 2.
 DR Pfam; PF00088; Trefoil; 1.
 DR SMART; SM00018; PD; 2.
 DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
 DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
 DR PROSITE; PS00025; P_TREFOIL; 1.
 KW Direct protein sequencing; Glycoprotein; Glycosidase; Hydrolase;
 KW Multifunctional enzyme; Repeat; Signal-anchor; Sulfation;
 KW Transmembrane.
 FT INIT_MET 0 0
 FT CHAIN 1 1840 Sucrase-isomaltase, intestinal.
 FT CHAIN 1 1012 Isomaltase.
 FT CHAIN 1013 1840 Sucrase.
 FT DOMAIN 1 11 Cytoplasmic.
 FT TRANSMEM 12 31 Signal-anchor for type II membrane
 FT protein.
 FT DOMAIN 32 1840 Luminal.
 FT DOMAIN 71 118 P-type.
 FT DOMAIN 119 1012 Isomaltase.
 FT DOMAIN 1013 1840 Sucrase.
 FT DOMAIN 42 69 Ser/Thr-rich.
 FT ACT_SITE 513 513 Nucleophile (By similarity).
 FT ACT_SITE 516 516 By similarity.
 FT ACT_SITE 1398 1398 Nucleophile (By similarity).
 FT ACT_SITE 1401 1401 By similarity.
 FT ACT_SITE 1511 1511 Proton donor (By similarity).
 FT DISULFID 72 103 By similarity.
 FT DISULFID 86 102 By similarity.
 FT DISULFID 97 115 By similarity.
 FT MOD_RES 400 400 Sulfotyrosine (Potential).
 FT MOD_RES 409 409 Sulfotyrosine (Potential).

FT MOD_RES 1386 1386 Sulfotyrosine (Potential).
 FT CARBOHYD 108 108 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 463 463 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 757 757 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 764 764 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 866 866 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 909 909 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1239 1239 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1307 1307 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1344 1344 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1358 1358 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1372 1372 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1484 1484 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1512 1512 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1574 1574 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1761 1761 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 1828 1828 N-linked (GlcNAc...) (Potential).
 FT CONFLICT 86 86 C -> W (in Ref. 1).
 FT CONFLICT 91 91 H -> S (in Ref. 2).
 FT CONFLICT 94 94 K -> Q (in Ref. 2).
 FT CONFLICT 222 222 C -> F (in Ref. 2).
 FT CONFLICT 230 230 L -> V (in Ref. 2).
 FT CONFLICT 240 240 T -> R (in Ref. 2).
 FT CONFLICT 252 252 G -> E (in Ref. 2).
 FT CONFLICT 283 283 L -> F (in Ref. 2).
 FT CONFLICT 294 294 G -> E (in Ref. 2).
 FT CONFLICT 326 326 T -> A (in Ref. 2).
 FT CONFLICT 350 352 YHW -> FNG (in Ref. 2).
 FT CONFLICT 359 361 S -> A (in Ref. 3).
 FT CONFLICT 732 732 E -> V (in Ref. 3).
 FT CONFLICT 735 735 E -> Q (in Ref. 3).
 FT CONFLICT 841 841 A -> T (in Ref. 3).
 FT CONFLICT 915 915 A -> R (in Ref. 3).
 FT CONFLICT 923 923 AG -> GT (in Ref. 3).
 FT CONFLICT 929 930 CR -> SQ (in Ref. 3).
 FT CONFLICT 937 938 GFCT -> ETDK (in Ref. 3).
 FT CONFLICT 958 961 Y -> C (in Ref. 3).
 FT CONFLICT 979 979 N -> H (in Ref. 3).
 FT CONFLICT 985 985 LP -> SL (in Ref. 3).
 FT CONFLICT 996 997 P -> A (in Ref. 3).
 FT CONFLICT 1009 1009 T -> P (in Ref. 3).
 FT CONFLICT 1022 1022 G -> E (in Ref. 3).
 FT CONFLICT 1026 1026 P -> K (in Ref. 3).
 FT CONFLICT 1031 1031 R -> S (in Ref. 3).
 FT CONFLICT 1093 1093 G -> A (in Ref. 3).
 FT CONFLICT 1098 1098 A -> D (in Ref. 3).
 FT CONFLICT 1301 1301 P -> A (in Ref. 3).
 FT CONFLICT 1336 1336 VW -> WG (in Ref. 3).
 FT CONFLICT 1338 1339 VW -> WG (in Ref. 3).
 SQ SEQUENCE 1840 AA; 210218 MW; 079AD45E9A23E2E CRC64;
 Query Match 27.8%; Score 1813.5; DB 1; Length 1840;
 Best Local Similarity 41.8%; Pred. No. 2.9e-95;
 Matches 378; Conservative 146; Mismatches 318; Indels 63; Gaps 20;
 QY 351 PRAVPTQCDVPNPSRFDCAPKAITQEQCEARGCCYIPAKOGLQAGQMGQPCFPFSPY 410
 DB 69 PGKCPPEQGEPINERINCIPQHPHTKAICEERGCGCWRPNWNTV-----IPWCFADNH- 121
 QY 411 SYKLENLSSEMGYTATLTR-TTTPFPFKDILTLDVMMETENRHLHTIKDPANRREY 469
 DB 122 GYNAESITNENAGLKATLNRIPTLPFGEDIKSVILTTQTGTGNFRFKITDPNNKREY 181
 QY 470 PLETRVHSRAPS--PLYSVFSFSEPPFGVIVHRQDLDGRVLLNTTVAFLPFADQFLQLSTS 527
 DB 182 PHQFVKETGTPAADTLVDVQVSENPFSIKVIRKSNKVLCDTSVGPLLYSNQYLQISTR 241
 QY 528 LPSQVITGLAHL-SPLMLSTSWTRITLWNRDLAP-TPCANLYGSHPFVLALED-GGSAH 584
 DB 242 LPSEVIYFGGHIHKRFRHDLYWKTWPIFTDRDEIPGDNHNLVGHQTFFMGIDTSGKSY 301
 QY 585 GVFLNSNMDVVLQPSALSWSRSTGGILDVYIFLGPEPKSVVQCYLDVGVGPPFPYWG 644
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 302 GVFLMNSNAMEVFIQTPITITRVTCGILDYFIPLGDTPEQVVOQVQVHWRPAMPAYVN 361
Qy 645 LGFHLRCWGSYSAITRQVVENMTRAHFLPDVQMNDDYMDRRDFTFNKQGRDPFPAWV 704
Db 362 LGFQLSRWYGSIDTVSEVVRNRREAGIPVDAQVTDIDIMEDHKFTYDRVFNGLPEFA 421
Qy 705 QELHQGRYMIWIDPAISSGPA--GSTRPVDEGLRGVFIYTNETG-QPLIGKWPWGST 761
Db 422 QDLNHHG-KYIILDPAISINKRANGAETQYVGRNEKQVWVNESDGTTFPLIGVWPGJT 480
Qy 762 APDPFTNPTALAWEDVABEFHDQVDFDGLWIDMNEPSNFIKSGED--GCPNNLENNPY 819
Db 481 VTPDFTNPOTIEWANECHLFHQQVBYDGLWIDMNEVSSFIQSLNKLGVLLIVLNPFP 540
Qy 820 VFGVVGTTLQAATICASSHQFLSTHYNLNHLXGLTEALASHRALVKA-RGTRPFPVSRST 878
Db 541 TFGILDKVMSKTLCDVAQHNGKQYDVHSLGYSMAIATEQAVERVFPNKRSLTRET 600
Qy 879 FAGHRYAGHTGVDVWSSWEQIASSVPEILQFNLLGVPLVGDVCGFLGNTSEELCVRT 938
Db 601 FCGSGRHHNLGIDNTASWEQMEWSITGMLEFGIFGMPVLVGATSCGFLADTTTELCCRWM 660
Qy 939 QLGAFYFPRNENSLSLPOEPVSEPAQAQMRKALTLRYALLPHLYTLFHOAHVAGET 998
Db 661 QLGAFYFPRNENNAEGYMQDPAYF---GQDSRHYLTIRYLLPPLYTLFRAHMFGET 717
Qy 999 VARPLFLEPPKDSSTVTDVHQLLWGBALLITPVLAGKAEVGTGYPFLGTWYDLQVPIEA 1058
Db 718 VARPLFLEYDNTSNWIEDTQFLMGALLITPVLRFGVENVSAYIPNATWYDET----- 772
Qy 1059 LGSLLPPPPAAPREPAIHSGQW-----VTLPAPLDTINVLHLAGYIIPLOGPGLTTTESRQ 1114
Db 773 -GIXRP-----WRKERINMYLPDGIHLRGYIIPTEQPDVTTTASRK 816
Qy 1115 QPMALAVALTKEGARGELFWDGSELEVLERCAYTQVIFLARNYVNLVRYVT-SEGA 1173
Db 817 NPLGLVLVLDNQAAKGELFWDGSKSIEKKYILYTFVSNLNLVNLCTHSSVAGET 876
Qy 1174 GLQLOKVTVLG-----VATAPQQLVSNGVPSNFTYSPDTKVLDT-CVSLLMGEQF 1223
Db 877 SLAFKTIKVLGUREDVRSITVGENDDQATH-----TNFTFDSANKILSITALNFNLAGSF 932
Qy 1224 LVSWC 1228
Db 933 IVRWC 937

RESULT 13
SUIS SUNMU
AC O62653, STANDARD, PRT; 1812 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Sucrase-isomaltase, intestinal [Contains: Sucrase (EC 3.2.1.48);
DE Isomaltase (EC 3.2.1.10)].
GN Name-SI:
OS Suncus murinus (House shrew) (Musk shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Crocidurinae; Suncus.
OX NCBI_TaxID=9378;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98298166; PubMed=9632713; DOI=10.1074/jbc.273.26.16464;
RA Ito T., Hayashi Y., Ohmori S., Oda S., Seo H.;
RT "Molecular cloning of sucrase-isomaltase cDNA in the house musk shrew
RT Suncus murinus and identification of a mutation responsible for
RT isolated sucrase deficiency.";
RL J. Biol. Chem. 273:16464-16469(1998).
CC -!- FUNCTION: Plays an important role in the final stage of
CC carbohydrate digestion (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of sucrose and maltose by an alpha-
CC D-glucosidase-type action.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,6-alpha-D-glucosidic linkages
CC in some oligosaccharides produced from starch and glycogen by
CC alpha-amylase, and in isomaltose.
CC -!- SUBUNIT: The resulting sucrose and isomaltase subunits stay
CC associated with one another in a complex by non-covalent linkages
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Brush border.
CC -!- PTM: The precursor is proteolytically cleaved when exposed to
CC pancreatic proteases in the intestinal lumen (By similarity).
CC -!- PTM: Sulfated (By similarity).
CC -!- MISCELLANEOUS: There is a high degree of homology between the
CC isomaltase and sucrose portions (41% of amino acid identity)
CC indicating that this protein is evolved by partial gene
CC duplication.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 31 family.
CC -!- SIMILARITY: Contains 1 P-type (trefoil) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011401; BAA25370.1; -;
DR InterPro; IPR000322; Glyco_hydro_31.
DR InterPro; IPR000519; P_trefoil.
DR Pfam; PF01055; Glyco_hydro_31; 2.
DR Pfam; PF00088; Trefoil; 1.
DR SMART; SM00018; PD; 2.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1; 2.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
DR PROSITE; PS00025; P_TREFOIL; 1.
KW Glycoprotein; Glycosidase; Hydrolase; Multifunctional enzyme; Repeat;
KW Signal-anchor; Sulfation; Transmembrane.
KW INIT MET 0 0 By similarity.
FT DOMAIN 1 11 Cytoplasmic (Potential).
FT DOMAIN 12 31 Signal-anchor for type II membrane
FT DOMAIN 12 31 protein (Potential).
FT FT Lumenal (Potential).
FT FT P-type.
FT FT 32 1812 Isomaltase.
FT FT 46 93 Sucrase.
FT FT 94 1812 Ser/Thr-rich.
FT FT 41 44 Nucleophile (By similarity).
FT FT 490 490 By similarity.
FT FT 493 493 Nucleophile (By similarity).
FT FT 1379 1379 By similarity.
FT FT 1382 1382 Proton donor (By similarity).
FT FT 1485 1485 By similarity.
FT FT 47 78 By similarity.
FT FT 61 77 By similarity.
FT FT 72 90 By similarity.
FT FT 376 376 Sulfotyrosine (Potential).
FT FT 1293 1293 Sulfotyrosine (Potential).
FT FT 1367 1367 Sulfotyrosine (Potential).
FT FT 1370 1370 Sulfotyrosine (Potential).
FT FT 126 126 N-linked (GlcNAc..) (Potential).
FT FT 387 387 N-linked (GlcNAc..) (Potential).
FT FT 668 668 N-linked (GlcNAc..) (Potential).
FT FT 790 790 N-linked (GlcNAc..) (Potential).
FT FT 895 895 N-linked (GlcNAc..) (Potential).
FT FT 910 910 N-linked (GlcNAc..) (Potential).
FT FT 1220 1220 N-linked (GlcNAc..) (Potential).
FT FT 1288 1288 N-linked (GlcNAc..) (Potential).
FT FT 1325 1325 N-linked (GlcNAc..) (Potential).
FT FT 1339 1339 N-linked (GlcNAc..) (Potential).
FT FT 1431 1431 N-linked (GlcNAc..) (Potential).
FT FT 1520 1520 N-linked (GlcNAc..) (Potential).
FT FT 1544 1544 N-linked (GlcNAc..) (Potential).
FT FT 1557 1557 N-linked (GlcNAc..) (Potential).
FT FT 1702 1702 N-linked (GlcNAc..) (Potential).
FT FT 1771 1771 N-linked (GlcNAc..) (Potential).

SQ	SEQUENCE	1812 AA; 208172 MW; 3ED93E407D783158 CRC64;	AMRP_HUMAN	STANDARD;	PRT; 357 AA.
Query Match	27.7%; Score 1806.5; DB 1; Length 1812;				
Best Local Similarity	40.4%; Pred. No. 71e-95;				
Matches 365; Conservative 163; Mismatches 315; Indels 61; Gaps 21;					
QY 354	VPTQCDVPNSRPD----CAPDKAITQEOCEARGCCYIPAKOGLQAGQWQCWCFPPPSY 409		DE	Alpha-2-macroglobulin receptor-associated protein precursor (Alpha-2-MRAP) (low density lipoprotein receptor-related protein-associated protein 1) (RAP).	
Db 43	IPGKCPAENDRLEKINCIPDQFPPTQALCAMQGCCWPNRNS-----PTWCSPANNH 96		GN	Name=LRPAP1; Synonyms=A2MRAP; Homo sapiens (Human)	
QY 410	PSYKLENLSSEMGYATITRTT-PTFFPKDILTLVDVMMETENRLHTIIXDPANRYE 468		OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Db 97	-GYEFKISNPINFPNLKNSPPTLFGDNTLNLLTTQSTANRFRKFTDPNNQRYE 155		OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
QY 469	VPLETPRVHSRAP--SPLYSEFESEPEFGVIVHROLDGRVLLNTTVAFLFADQPLQLS 525		OC	NCBI_TaxID=9606;	
Db 156	VHOFVKNKDFGPPASNPIDYKITEPFSIKVIRKSNKKILFDTSIGPLVSNQVLQIS 215		OC	[1]	SEQUENCE FROM N.A., AND SEQUENCE OF 183-195 AND 257-272.
QY 526	TSLPSQYITGLAEHL-SPLMLSTSWTRITLWNRDLAP-TPGANLYGSHPHYLALED-GGS 582		OC	TISSUE=Placenta;	
Db 216	TKLPSKYIYGLGEVHKRFRHDLYNKTFIFTRDQLPGDNNNNLYGHQTFWMSIEDTSK 275		OC	MEDLINE=91302371; PubMed=1712782;	
QY 583	ANGVFLNNSAMDVLOPSALSWRSTGGILDVYIFELGPEKSVVQOYLDVGYFPMPPY 642		OC	Strickland D.K., Ashcom J.D., Williams S., Battey F., Behre E.,	
Db 276	SFGVFLMNSNAMEVFIQPTPIVTVIGILDYIFLGDTPGQVQOYQELTGRPAMPY 335		OC	McTigue K., Battey J.F., Argraves W.S.;	
QY 643	WGLGFHLCKWGSSTAITQVVENMTRAHPFLDQVNDLDYMDSRDFTFNKDGFRDPA 702		OC	"Primary structure of alpha 2-macroglobulin receptor-associated protein. Human homologue of a Heymann nephritis antigen.";	
Db 336	WSLGFQLSRNYSGLDAVKEVVKENRDARIIPDAQVTDIDYMEKDKFTYNNKTFYGLPE 395		OC	J. Biol. Chem. 266:13364-13369(1991).	
QY 703	MYQELHQGRRYMMIVDPAISSGPGAS--YRPYDEGLRRGVFITNETG-QPLIGKWPFG 759		OC	[2]	SEQUENCE FROM N.A., AND VARIANT MET-311.
Db 396	FVKDLHDHGOKYIILDPALISITSLANGNHYKTYERQEKVWYQSDGTTLLGEVWPG 455		OC	MEDLINE=99008833; PubMed=9782079; DOI=10.1006/geno.1998.5413;	
QY 760	STAFPDFTPTALAWMEDVAEFHDQVDPGLDWMDEPSNTRGSDCCPNNELENPPY 819		OC	Van Leuven F., Thiry E., Stas L., Nelissen B.;	
Db 456	LTVPDFTNPKCDMTNCSIFHEIKYDGLWIDMNEVSFVHGTSKCSNKLNYPPF 515		OC	"Analysis of the human LRPAP1 gene coding for the lipoprotein receptor-associated protein: identification of 22 polymorphisms and one mutation.";	
QY 820	VPQVGGTLOAATICASSHQFLSTHNLNLYGLTEAISHALVKA-RGTGPPFVISRT 878		OC	Genomics 52:145-151(1998).	
Db 516	IPDILDKMYAKTICMDAIOHWGKQYDVHSLGYSAIATEKAEKVPFNKSFILTRST 575		OC	[3]	SEQUENCE FROM N.A.
QY 879	FAGHGYAGHTGDVWSSWEQLASSVPEILQFNLLGVPLVGDVCGFLGNTSEELCVRT 938		OC	Hunt A.;	
Db 576	FAGTGKHATHLGDNTPSWEHMEWSITPMLFGLGMPFAGIDICGVVDVTTTELCTRW 635		OC	Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.	
QY 939	QLGAFYFPRNHNLSLPOEYSPSEPA--QAMRKALTLRYALLPHLYTLFHOAHVAG 996		OC	[4]	SEQUENCE OF 204-250 FROM N.A.
Db 636	QIGAFYFPRDNAGGYMPQDPAYFGQDLSLLVNTSRHYLDIWTLLPYLLNLYKAYVYG 695		OC	MEDLINE=95309917; PubMed=7789983;	
QY 997	ETVARPLFEPKDSSTWTDHQLWGEALLITPVLOAGKAEVTVGYPLGTWYDLQTVPI 1056		OC	Van Leuven F., Hilliker C., Sernels L., Umans L., Overbergh L.,	
Db 696	ETVARPLFVEFDNTSWIEDQLWGSALLITPVLRQADRMSAYIPDATWYDET--- 752		OC	Strooper B., Frys J.-P., Van den Berghe H.;	
QY 1057	EALGLSPPPPAAPREPAIHSEQW----VTLPAPLDTINVHLRAGYIIPLOGPGHTTSES 1112		OC	"Cloning, characterization, and chromosomal localization to 4p16 of the human gene (LRPAP1) coding for the alpha 2-macroglobulin receptor-associated protein and structural comparison with the murine gene coding for the 44-kDa heparin-binding protein.";	
Db 753	-----GKRTWRKQRMVYLPBGKIGLHVRGGYIIPQCPAVNTTAS 794		OC	Genomics 25:492-500(1995).	
QY 1113	RQPMALAVATKGGEARGELFWDDGESLEVLERGAYTOVIFLARNTIVNELVRVS--- 1170		OC	[5]	CHARACTERIZATION.
Db 795	RKNPLGLIILADNNA-AKGDFPDDGSKDSIEKGYILTYFSLVNLNEL--DIICHTSY 851		OC	MEDLINE=93016038; PubMed=1400426;	
QY 1171	-EGAGLOKQVTVLGA-TPAQ-QVLSNG---VPVSNFTYSPTDKVLDI-CVSLMAGEOF 1223		OC	Kounnas M.Z., Argraves W.S., Strickland D.K.;	
Db 852	QEGTTLAFETIKILGLANTVTVQVANNQOQIIHNSFTYHASNQSLLIDNLKLNKGNP 911		OC	"The 39-kDa receptor-associated protein interacts with two members of the low density lipoprotein receptor family, alpha 2-macroglobulin receptor and glycoprotein 330.";	
QY 1224	LVSQ 1227		OC	J. Biol. Chem. 267:21162-21166(1992).	
Db 912	TVQW 915		OC	[6]	SUBCELLULAR LOCATION, AND INTERACTION WITH LRP1.
			OC	Bu G., Geuze H.J., Strous G.J., Schwartz A.L.;	
			OC	"39 kDa receptor-associated protein is an ER resident protein and molecular chaperone for LDL receptor-related protein.";	
			OC	EMBO J. 14:2269-2280(1995).	
			OC	[7]	STRUCTURE BY NMR OF 51-131.
			OC	MEDLINE=97352831; PubMed=9207124; DOI=10.1073/pnas.94.14.7521;	
			OC	Nielsen P.R., Ellgaard L., Etzerodt M., Thøgersen H.C., Poulsen F.M.;	
			OC	"The solution structure of the N-terminal domain of alpha-2-macroglobulin receptor-associated protein.";	
			OC	Proc. Natl. Acad. Sci. U.S.A. 94:7521-7525(1997).	
			OC	-/- FUNCTION: Interacts with LRP1/alpha-2-macroglobulin receptor and glycoprotein 330.	
			OC	-/- SUBUNIT: Present on cell surface forming a complex with the alpha-2-macroglobulin receptor heavy and light chains.	
			OC	-/- SUBCELLULAR LOCATION: Intracellular and associated with cell surface receptors. Found in the endoplasmic reticulum.	
			OC	-/- DISEASE: In complex with the alpha-2-MR or gp330, it may have some	

role in the pathogenesis of membrane glomerular nephritis.
-1- SIMILARITY: Belongs to the alpha-2-MBP family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M63959; AAA51553.1; -;
DR EMBL; AF033767; AAC67373.1; -;
DR EMBL; AF033760; AAC67373.1; JOINED.
DR EMBL; AF033761; AAC67373.1; JOINED.
DR EMBL; AF033762; AAC67373.1; JOINED.
DR EMBL; AF033763; AAC67373.1; JOINED.
DR EMBL; AF033764; AAC67373.1; JOINED.
DR EMBL; AF033765; AAC67373.1; JOINED.
DR EMBL; AF033766; AAC67373.1; JOINED.
DR EMBL; Z69710; -; NOT ANNOTATED_CDS.
DR EMBL; U06976; AAA87889.1; -;
DR PIR; A39875; A39875.
DR PDB; 1LRE; NMR; @=51-131.
DR PDB; 1NRE; NMR; @=51-131.
DR Genew; HGNC:6701; LRPA1.
DR MIM; 104225; -;
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0005886; C:plasma membrane; TAS.
DR GO; GO:0004873; F:asialoglycoprotein receptor activity; TAS.
DR GO; GO:0005509; F:calcium ion binding; TAS.
DR GO; GO:0003754; F:chaperone activity; TAS.
DR GO; GO:0008201; F:heparin binding; TAS.
DR GO; GO:0008034; F:cell proliferation; TAS.
DR GO; GO:000457; P:protein folding; TAS.
DR GO; GO:0016192; P:vesicle-mediated transport; TAS.
DR InterPro; IPR010483; Alpha-2-MBP_C.
DR InterPro; IPR000886; ER target_S.
DR InterPro; IPR009066; MGRcppt_assoc_1.
DR Pfam; PF06401; Alpha-2-MBP_C; 1.
DR Pfam; PF06400; Alpha-2-MBP_N; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR 3D-structure; Antigen; Direct protein sequencing;
KW Endoplasmic reticulum; Glycoprotein; Heparin-binding; Polymorphism;
KW Signal.
FT SIGNAL 1 34 Potential.
FT CHAIN 35 357 Alpha-2-macroglobulin receptor-associated
protein.
FT DOMAIN 237 353 LDL receptor binding (Potential).
FT CARBOHYD 268 268 N-linked (GlcNAc...) (Potential).
FT SITE 354 357 Prevent secretion from ER (Potential).
FT VARIANT 311 311 V -> M (in dbSNP:1800493).
FT FTID=VAR_011821.
FT HELIX 57 68
FT TURN 69 70
FT HELIX 73 99
FT TURN 100 100
FT TURN 103 104
FT TURN 107 122
FT HELIX 123 123
FT TURN 124 124
FT TURN 125 125
FT TURN 126 126
FT TURN 127 127
FT TURN 128 128
FT TURN 129 129
FT TURN 130 130
FT TURN 131 131
FT TURN 132 132
FT TURN 133 133
FT TURN 134 134
FT TURN 135 135
FT TURN 136 136
FT TURN 137 137
FT TURN 138 138
FT TURN 139 139
FT TURN 140 140
FT TURN 141 141
FT TURN 142 142
FT TURN 143 143
FT TURN 144 144
FT TURN 145 145
FT TURN 146 146
FT TURN 147 147
FT TURN 148 148
FT TURN 149 149
FT TURN 150 150
FT TURN 151 151
FT TURN 152 152
FT TURN 153 153
FT TURN 154 154
FT TURN 155 155
FT TURN 156 156
FT TURN 157 157
FT TURN 158 158
FT TURN 159 159
FT TURN 160 160
FT TURN 161 161
FT TURN 162 162
FT TURN 163 163
FT TURN 164 164
FT TURN 165 165
FT TURN 166 166
FT TURN 167 167
FT TURN 168 168
FT TURN 169 169
FT TURN 170 170
FT TURN 171 171
FT TURN 172 172
FT TURN 173 173
FT TURN 174 174
FT TURN 175 175
FT TURN 176 176
FT TURN 177 177
FT TURN 178 178
FT TURN 179 179
FT TURN 180 180
FT TURN 181 181
FT TURN 182 182
FT TURN 183 183
FT TURN 184 184
FT TURN 185 185
FT TURN 186 186
FT TURN 187 187
FT TURN 188 188
FT TURN 189 189
FT TURN 190 190
FT TURN 191 191
FT TURN 192 192
FT TURN 193 193
FT TURN 194 194
FT TURN 195 195
FT TURN 196 196
FT TURN 197 197
FT TURN 198 198
FT TURN 199 199
FT TURN 200 200
FT TURN 201 201
FT TURN 202 202
FT TURN 203 203
FT TURN 204 204
FT TURN 205 205
FT TURN 206 206
FT TURN 207 207
FT TURN 208 208
FT TURN 209 209
FT TURN 210 210
FT TURN 211 211
FT TURN 212 212
FT TURN 213 213
FT TURN 214 214
FT TURN 215 215
FT TURN 216 216
FT TURN 217 217
FT TURN 218 218
FT TURN 219 219
FT TURN 220 220
FT TURN 221 221
FT TURN 222 222
FT TURN 223 223
FT TURN 224 224
FT TURN 225 225
FT TURN 226 226
FT TURN 227 227
FT TURN 228 228
FT TURN 229 229
FT TURN 230 230
FT TURN 231 231
FT TURN 232 232
FT TURN 233 233
FT TURN 234 234
FT TURN 235 235
FT TURN 236 236
FT TURN 237 237
FT TURN 238 238
FT TURN 239 239
FT TURN 240 240
FT TURN 241 241
FT TURN 242 242
FT TURN 243 243
FT TURN 244 244
FT TURN 245 245
FT TURN 246 246
FT TURN 247 247
FT TURN 248 248
FT TURN 249 249
FT TURN 250 250
FT TURN 251 251
FT TURN 252 252
FT TURN 253 253
FT TURN 254 254
FT TURN 255 255
FT TURN 256 256
FT TURN 257 257
FT TURN 258 258
FT TURN 259 259
FT TURN 260 260
FT TURN 261 261
FT TURN 262 262
FT TURN 263 263
FT TURN 264 264
FT TURN 265 265
FT TURN 266 266
FT TURN 267 267
FT TURN 268 268
FT TURN 269 269
FT TURN 270 270
FT TURN 271 271
FT TURN 272 272
FT TURN 273 273
FT TURN 274 274
FT TURN 275 275
FT TURN 276 276
FT TURN 277 277
FT TURN 278 278
FT TURN 279 279
FT TURN 280 280
FT TURN 281 281
FT TURN 282 282
FT TURN 283 283
FT TURN 284 284
FT TURN 285 285
FT TURN 286 286
FT TURN 287 287
FT TURN 288 288
FT TURN 289 289
FT TURN 290 290
FT TURN 291 291
FT TURN 292 292
FT TURN 293 293
FT TURN 294 294
FT TURN 295 295
FT TURN 296 296
FT TURN 297 297
FT TURN 298 298
FT TURN 299 299
FT TURN 300 300
FT TURN 301 301
FT TURN 302 302
FT TURN 303 303
FT TURN 304 304
FT TURN 305 305
FT TURN 306 306
FT TURN 307 307
FT TURN 308 308
FT TURN 309 309
FT TURN 310 310
FT TURN 311 311
FT TURN 312 312
FT TURN 313 313
FT TURN 314 314
FT TURN 315 315
FT TURN 316 316
FT TURN 317 317
FT TURN 318 318
FT TURN 319 319
FT TURN 320 320
FT TURN 321 321
FT TURN 322 322
FT TURN 323 323
FT TURN 324 324
FT TURN 325 325
FT TURN 326 326
FT TURN 327 327
FT TURN 328 328
FT TURN 329 329
FT TURN 330 330
FT TURN 331 331
FT TURN 332 332
FT TURN 333 333
FT TURN 334 334
FT TURN 335 335
FT TURN 336 336
FT TURN 337 337
FT TURN 338 338
FT TURN 339 339
FT TURN 340 340
FT TURN 341 341
FT TURN 342 342
FT TURN 343 343
FT TURN 344 344
FT TURN 345 345
FT TURN 346 346
FT TURN 347 347
FT TURN 348 348
FT TURN 349 349
FT TURN 350 350
FT TURN 351 351
FT TURN 352 352
FT TURN 353 353
FT TURN 354 354
FT TURN 355 355
FT TURN 356 356
FT TURN 357 357
FT TURN 358 358
FT TURN 359 359
FT TURN 360 360
FT TURN 361 361
FT TURN 362 362
FT TURN 363 363
FT TURN 364 364
FT TURN 365 365
FT TURN 366 366
FT TURN 367 367
FT TURN 368 368
FT TURN 369 369
FT TURN 370 370
FT TURN 371 371
FT TURN 372 372
FT TURN 373 373
FT TURN 374 374
FT TURN 375 375
FT TURN 376 376
FT TURN 377 377
FT TURN 378 378
FT TURN 379 379
FT TURN 380 380
FT TURN 381 381
FT TURN 382 382
FT TURN 383 383
FT TURN 384 384
FT TURN 385 385
FT TURN 386 386
FT TURN 387 387
FT TURN 388 388
FT TURN 389 389
FT TURN 390 390
FT TURN 391 391
FT TURN 392 392
FT TURN 393 393
FT TURN 394 394
FT TURN 395 395
FT TURN 396 396
FT TURN 397 397
FT TURN 398 398
FT TURN 399 399
FT TURN 400 400
FT TURN 401 401
FT TURN 402 402
FT TURN 403 403
FT TURN 404 404
FT TURN 405 405
FT TURN 406 406
FT TURN 407 407
FT TURN 408 408
FT TURN 409 409
FT TURN 410 410
FT TURN 411 411
FT TURN 412 412
FT TURN 413 413
FT TURN 414 414
FT TURN 415 415
FT TURN 416 416
FT TURN 417 417
FT TURN 418 418
FT TURN 419 419
FT TURN 420 420
FT TURN 421 421
FT TURN 422 422
FT TURN 423 423
FT TURN 424 424
FT TURN 425 425
FT TURN 426 426
FT TURN 427 427
FT TURN 428 428
FT TURN 429 429
FT TURN 430 430
FT TURN 431 431
FT TURN 432 432
FT TURN 433 433
FT TURN 434 434
FT TURN 435 435
FT TURN 436 436
FT TURN 437 437
FT TURN 438 438
FT TURN 439 439
FT TURN 440 440
FT TURN 441 441
FT TURN 442 442
FT TURN 443 443
FT TURN 444 444
FT TURN 445 445
FT TURN 446 446
FT TURN 447 447
FT TURN 448 448
FT TURN 449 449
FT TURN 450 450
FT TURN 451 451
FT TURN 452 452
FT TURN 453 453
FT TURN 454 454
FT TURN 455 455
FT TURN 456 456
FT TURN 457 457
FT TURN 458 458
FT TURN 459 459
FT TURN 460 460
FT TURN 461 461
FT TURN 462 462
FT TURN 463 463
FT TURN 464 464
FT TURN 465 465
FT TURN 466 466
FT TURN 467 467
FT TURN 468 468
FT TURN 469 469
FT TURN 470 470
FT TURN 471 471
FT TURN 472 472
FT TURN 473 473
FT TURN 474 474
FT TURN 475 475
FT TURN 476 476
FT TURN 477 477
FT TURN 478 478
FT TURN 479 479
FT TURN 480 480
FT TURN 481 481
FT TURN 482 482
FT TURN 483 483
FT TURN 484 484
FT TURN 485 485
FT TURN 486 486
FT TURN 487 487
FT TURN 488 488
FT TURN 489 489
FT TURN 490 490
FT TURN 491 491
FT TURN 492 492
FT TURN 493 493
FT TURN 494 494
FT TURN 495 495
FT TURN 496 496
FT TURN 497 497
FT TURN 498 498
FT TURN 499 499
FT TURN 500 500
FT TURN 501 501
FT TURN 502 502
FT TURN 503 503
FT TURN 504 504
FT TURN 505 505
FT TURN 506 506
FT TURN 507 507
FT TURN 508 508
FT TURN 509 509
FT TURN 510 510
FT TURN 511 511
FT TURN 512 512
FT TURN 513 513
FT TURN 514 514
FT TURN 515 515
FT TURN 516 516
FT TURN 517 517
FT TURN 518 518
FT TURN 519 519
FT TURN 520 520
FT TURN 521 521
FT TURN 522 522
FT TURN 523 523
FT TURN 524 524
FT TURN 525 525
FT TURN 526 526
FT TURN 527 527
FT TURN 528 528
FT TURN 529 529
FT TURN 530 530
FT TURN 531 531
FT TURN 532 532
FT TURN 533 533
FT TURN 534 534
FT TURN 535 535
FT TURN 536 536
FT TURN 537 537
FT TURN 538 538
FT TURN 539 539
FT TURN 540 540
FT TURN 541 541
FT TURN 542 542
FT TURN 543 543
FT TURN 544 544
FT TURN 545 545
FT TURN 546 546
FT TURN 547 547
FT TURN 548 548
FT TURN 549 549
FT TURN 550 550
FT TURN 551 551
FT TURN 552 552
FT TURN 553 553
FT TURN 554 554
FT TURN 555 555
FT TURN 556 556
FT TURN 557 557
FT TURN 558 558
FT TURN 559 559
FT TURN 560 560
FT TURN 561 561
FT TURN 562 562
FT TURN 563 563
FT TURN 564 564
FT TURN 565 565
FT TURN 566 566
FT TURN 567 567
FT TURN 568 568
FT TURN 569 569
FT TURN 570 570
FT TURN 571 571
FT TURN 572 572
FT TURN 573 573
FT TURN 574 574
FT TURN 575 575
FT TURN 576 576
FT TURN 577 577
FT TURN 578 578
FT TURN 579 579
FT TURN 580 580
FT TURN 581 581
FT TURN 582 582
FT TURN 583 583
FT TURN 584 584
FT TURN 585 585
FT TURN 586 586
FT TURN 587 587
FT TURN 588 588
FT TURN 589 589
FT TURN 590 590
FT TURN 591 591
FT TURN 592 592
FT TURN 593 593
FT TURN 594 594
FT TURN 595 595
FT TURN 596 596
FT TURN 597 597
FT TURN 598 598
FT TURN 599 599
FT TURN 600 600
FT TURN 601 601
FT TURN 602 602
FT TURN 603 603
FT TURN 604 604
FT TURN 605 605
FT TURN 606 606
FT TURN 607 607
FT TURN 608 608
FT TURN 609 609
FT TURN 610 610
FT TURN 611 611
FT TURN 612 612
FT TURN 613 613
FT TURN 614 614
FT TURN 615 615
FT TURN 616 616
FT TURN 617 617
FT TURN 618 618
FT TURN 619 619
FT TURN 620 620
FT TURN 621 621
FT TURN 622 622
FT TURN 623 623
FT TURN 624 624
FT TURN 625 625
FT TURN 626 626
FT TURN 627 627
FT TURN 628 628
FT TURN 629 629
FT TURN 630 630
FT TURN 631 631
FT TURN 632 632
FT TURN 633 633
FT TURN 634 634
FT TURN 635 635
FT TURN 636 636
FT TURN 637 637
FT TURN 638 638
FT TURN 639 639
FT TURN 640 640
FT TURN 641 641
FT TURN 642 642
FT TURN 643 643
FT TURN 644 644
FT TURN 645 645
FT TURN 646 646
FT TURN 647 647
FT TURN 648 648
FT TURN 649 649
FT TURN 650 650
FT TURN 651 651
FT TURN 652 652
FT TURN 653 653
FT TURN 654 654
FT TURN 655 655
FT TURN 656 656
FT TURN 657 657
FT TURN 658 658
FT TURN 659 659
FT TURN 660 660
FT TURN 661 661
FT TURN 662 662
FT TURN 663 663
FT TURN 664 664
FT TURN 665 665
FT TURN 666 666
FT TURN 667 667
FT TURN 668 668
FT TURN 669 669
FT TURN 670 670
FT TURN 671 671
FT TURN 672 672
FT TURN 673 673
FT TURN 674 674
FT TURN 675 675
FT TURN 676 676
FT TURN 677 677
FT TURN 678 678
FT TURN 679 679
FT TURN 680 680
FT TURN 681 681
FT TURN 682 682
FT TURN 683 683
FT TURN 684 684
FT TURN 685 685
FT TURN 686 686
FT TURN 687 687
FT TURN 688 688
FT TURN 689 689
FT TURN 690 690
FT TURN 691 691
FT TURN 692 692
FT TURN 693 693
FT TURN 694 694
FT TURN 695 695
FT TURN 696 696
FT TURN 697 697
FT TURN 698 698
FT TURN 699 699
FT TURN 700 700
FT TURN 701 701
FT TURN 702 702
FT TURN 703 703
FT TURN 704 704
FT TURN 705 705
FT TURN 706 706
FT TURN 707 707
FT TURN 708 708
FT TURN 709 709
FT TURN 710 710
FT TURN 711 711
FT TURN 712 712
FT TURN 713 713
FT TURN 714 714
FT TURN 715 715
FT TURN 716 716
FT TURN 717 717
FT TURN 718 718
FT TURN 719 719
FT TURN 720 720
FT TURN 721 721
FT TURN 722 722
FT TURN 723 723
FT TURN 724 724
FT TURN 725 725
FT TURN 726 726
FT TURN 727 727
FT TURN 728 728
FT TURN 729 729
FT TURN 730 730
FT TURN 731 731
FT TURN 732 732
FT TURN 733 733
FT TURN 734 734
FT TURN 735 735
FT TURN 736 736
FT TURN 737 737
FT TURN 738 738
FT TURN 739 739
FT TURN 740 740
FT TURN 741 741
FT TURN 742 742
FT TURN 743 743
FT TURN 744 744
FT TURN 745 745
FT TURN 746 746
FT TURN 747 747
FT TURN 748 748
FT TURN 749 749
FT TURN 750 750
FT TURN 751 751
FT TURN 752 752
FT TURN 753 753
FT TURN 754 754
FT TURN 755 755
FT TURN 756 756
FT TURN 757 757
FT TURN 758 758
FT TURN 759 759
FT TURN 760 760
FT TURN 761 761
FT TURN 762 762
FT TURN 763 763
FT TURN 764 764
FT TURN 765 765
FT TURN 766 766
FT TURN 767 767
FT TURN 768 768
FT TURN 769 769
FT TURN 770 770
FT TURN 771 771
FT TURN 772 772
FT TURN 773 773
FT TURN 774 774
FT TURN 775 775
FT TURN 776 776
FT TURN 777 777
FT TURN 778 778
FT TURN 779 779
FT TURN 780 780
FT TURN 781 781
FT TURN 782 782
FT TURN 783 783
FT TURN 784 784
FT TURN 785 785
FT TURN 786 786
FT TURN 787 787
FT TURN 788 788
FT TURN 789 789
FT TURN 790 790
FT TURN 791 791
FT TURN 792 792
FT TURN 793 793
FT TURN 794 794
FT TURN 795 795
FT TURN 796 796
FT TURN 797 797
FT TURN 798 798
FT TURN 799 799
FT TURN 800 800
FT TURN 801 801
FT TURN 802 802
FT TURN 803 803
FT TURN 804 804
FT TURN 805 805
FT TURN 806 806
FT TURN 807 807
FT TURN 808 808
FT TURN 809 809
FT TURN 810 810
FT TURN 811 811
FT TURN 812 812
FT TURN 813 813
FT TURN 814 814
FT TURN 815 815
FT TURN 816 816
FT TURN 817 817
FT TURN 818 818
FT TURN 819 819
FT TURN 820 820
FT TURN 821 821
FT TURN 822 822
FT TURN 823 823
FT TURN 824 824
FT TURN 825 825
FT TURN 826 826
FT TURN 827 827
FT TURN 828 828
FT TURN 829 829
FT TURN 830 830
FT TURN 831 831
FT TURN 832 832
FT TURN 833 833
FT TURN 834 834
FT TURN 835 835
FT TURN 836 836
FT TURN 837 837
FT TURN 838 838
FT TURN 839 839
FT TURN 840 840
FT TURN 841 841
FT TURN 842 842
FT TURN 843 843
FT TURN 844 844
FT TURN 845 845
FT TURN 846 846
FT TURN 847 847
FT TURN 848 848
FT TURN 849 849
FT TURN 850 850
FT TURN 851 851
FT TURN 852 852
FT TURN 853 853
FT TURN 854 854
FT TURN 855 855
FT TURN 856 856
FT TURN 857 857
FT TURN 858 858
FT TURN 859 859
FT TURN 860 860
FT TURN 861 861
FT TURN 862 862
FT TURN 863 863
FT TURN 864 864
FT TURN 865 865
FT TURN 866 866
FT TURN 867 867
FT TURN 868 868
FT TURN 869 869
FT TURN 870 870
FT TURN 871 871
FT TURN 872 872
FT TURN 873 873
FT TURN 874 874
FT TURN 875 875
FT TURN 876 876
FT TURN 877 877
FT TURN 878 878
FT TURN 879 879
FT TURN 880 880
FT TURN 881 881
FT TURN 882 882
FT TURN 883 883
FT TURN 884 884
FT TURN 885 885
FT TURN 886 886
FT TURN 887 887
FT TURN 888 888
FT TURN 889 889
FT TURN 890 890
FT TURN 891 891
FT TURN 892 892
FT TURN 893 893
FT TURN 894 894
FT TURN 895 895
FT TURN 896 896
FT TURN 897 897
FT TURN 898 898
FT TURN 899 899
FT TURN 900 900
FT TURN 901 901
FT TURN 902 902
FT TURN 903 903
FT TURN 904 904
FT TURN 905 905
FT TURN 906 906
FT TURN 907 907
FT TURN 908 908
FT TURN 909 909
FT TURN 910 910
FT TURN 911 911
FT TURN 912 912
FT TURN 913 913
FT TURN 914 914
FT TURN 915 915
FT TURN 916 916
FT TURN 917 917
FT TURN 918 918
FT TURN 919 919
FT TURN 920 920
FT TURN 921 921
FT TURN 922 922
FT TURN 923 923
FT TURN 924 924
FT TURN 925 925
FT TURN 926 926
FT TURN 927 927
FT TURN 928 928
FT TURN 929 929
FT TURN 930 930
FT TURN 931 931
FT TURN 932 932
FT TURN 933 933
FT TURN 934 934
FT TURN 935 935
FT TURN 936 936
FT TURN 937 937
FT TURN 938 938
FT TURN 939 939
FT TURN 940 940
FT TURN 941 941
FT TURN 942 942
FT TURN 943 943
FT TURN 944 944
FT TURN 945 945
FT TURN 946 946
FT TURN 947 947
FT TURN 948 948
FT TURN 949 949
FT TURN 950 950
FT TURN 951 951
FT TURN 952 952
FT TURN 953 953
FT TURN 954 954
FT TURN 955 955
FT TURN 956 956
FT TURN 957 957
FT TURN 958 958
FT TURN 959 959
FT TURN 960 960
FT TURN 961 961
FT TURN 962 962
FT TURN 963 963
FT TURN 964 964
FT TURN 965 965
FT TURN 966 966
FT TURN 967 967
FT TURN 968 968
FT TURN 969 969
FT TURN 970 970
FT TURN 971 971
FT TURN 972 972
FT TURN 973 973
FT TURN 974 974
FT TURN 975 975
FT TURN 976 976
FT TURN 977 977
FT TURN 978 978
FT TURN 979 979
FT TURN 980 980
FT TURN 981 981
FT TURN 982 982
FT TURN 983 983
FT TURN 984 984
FT TURN 985 985
FT TURN 986 986
FT TURN 987 987
FT TURN 988 988
FT TURN 989 989
FT TURN 990 990
FT TURN 991 991
FT TURN 992 992
FT TURN 993 993
FT TURN 994 994
FT TURN 995 995
FT TURN 996 996
FT TURN 997 997
FT TURN 998 998
FT TURN 999 999
FT TURN 1000 1000
FT TURN 1001 1001
FT TURN 1002 1002
FT TURN 1003 1003
FT TURN 1004 1004
FT TURN 1005 1005
FT TURN 1006 1006
FT TURN 1007 1007
FT TURN 1008 1008
FT TURN 1009 1009
FT TURN 1010 1010
FT TURN 1011 1011
FT TURN 1012 1012
FT TURN 1013 1013
FT TURN 1014 1014
FT TURN 1015 1015
FT TURN 1016 1016
FT TURN 1017 1017
FT TURN 1018 1018
FT TURN 1019 1019
FT TURN 1020 1020
FT TURN 1021 1021
FT TURN 1022 1022
FT TURN 1023 1023
FT TURN 1024 1024
FT TURN 1025 1025
FT TURN 1026 1026
FT TURN 1027 1027
FT TURN 1028 1028
FT TURN 1029 1029
FT TURN 1030 1030
FT TURN 1031 1031
FT TURN 1032 1032
FT TURN 1033 1033
FT TURN

Db 299 GYKQVSDITNVVNNYKNSQIPLDVWDDDDHMDGAKDFTLDPINYPEYKLRPFLORIHAN 358
QY 711 GRZYMIVDPAISSGPAGSYREYDEGLARGVPIITNETQPLIGKVMRGSTAFPPDTNET 770
Db 359 GMYVVLIIIDPGIAIN---TSYGTFRQGMADDVFI-KHGGSPFLGQWPGAVYFPDFLNP 414
QY 771 ALAWNEDMVAEFHQVFPDGLWIDMNEPSNFIIRG-----SEDCPN----- 811
Db 415 TVNFWADEISHFHQWVPVDGLWIDMNEISNFCSGKCSIPTNRSCTGCTGPFWECCLDRTNI 474
QY 812 --NELNPPYVPGVGG--TLQAATTCASHOFLSTHYN-----LHNYGLTEAIAASHR 861
Db 475 TATRDVPPYKINASGTQVPLGFKTIATS-----SVHYNGVLEYDAHSYGLSQAIAIATHK 529
QY 862 ALVKARGTRPFIYSRSTFAGHGYAGHWTGCDYWSWEOQLASSVPEILOFNLLGVPLVGAD 921
Db 530 ALQNLLDKRPPVLTRETFVSGSGSYAAHWTGDNKATWEDLRYISITILNFMFGMPMVGAD 589
QY 922 VCGFLGNTSEELCVRWTLQCAFYPFMRNHNLSLSPQEPYFSEPAQQAAMRKALTLRYAL 981
Db 590 ICGFYDITTEELCGRWIQLGAFYFSDHNSLASKQEYLYLWDSVAKSA-RKALGLGYEL 648
QY 982 LPHLYTLFHOAHVAGETVARPLFLEPPKDSSTWVDHQLLMEALLITPVLOAGKAEVTG 1041
Db 649 LPYLYTLNYDAHTTGAPIARPLFFSPQDPETVAVSKQFLGPGVLIISPLYNKTTSVNA 708
QY 1042 YPPLGTWYDLQVPIEALGSLPPPPAAPREPAIHSEGOVTLPAPLDTINVHLRAGYIIP 1101
Db 709 YFPKGSWYNLNDMTM-----AVKSSGOYVTLQAPMDTINHVHCEGMILP 752
QY 1102 LQOPGLTTTESRQOPMALAVAL-----TKGGEARGELFWDGSELEV-LERGAYTQVIFL 1155
Db 753 MQRGMTTIVARWTPFTLIIAPLPGQSTGGKAKGHLFLDSGEDVDMKTAEGKSTYVDFS 812
QY 1156 ARNTIVNELVRVTSIGA-----GLQKQVTVLGVATA--POQVL--SNGVPV--SNFTY 1204
Db 813 AESDGKKVRLVSQVESGSYGLSQGWVVEKLMILGLSKSHLSQIAFQLDGKPFSSFTY 872
QY 1205 S 1205
Db 873 S 873

Search completed: April 5, 2005, 01:05:03
Job time : 155 secs